**Complete mitochondrial genome of *Camponotus atrox*
(Hymenoptera: Formicidae): a new tRNA arrangement in
Hymenoptera**

Journal:	<i>Genome</i>
Manuscript ID	gen-2015-0080.R1
Manuscript Type:	Article
Date Submitted by the Author:	03-Oct-2015
Complete List of Authors:	Kim, Min Jee; Chonnam National University, Applied Biology Hong, Eui Jeong; Korea National Park Service, National Park Research Institute Kim, Iksoo; Chonnam National University,
Keyword:	mitochondrial DNA, mt genome, Hymenoptera, ants, gene rearrangement

Complete mitochondrial genome of *Camponotus atrox* (Hymenoptera: Formicidae): a new tRNA arrangement in Hymenoptera

Min Jee Kim, Eui Jeong Hong and Iksoo Kim

Min Jee Kim and Iksoo Kim. College of Agriculture & Life Sciences, Chonnam National University, Gwangju 500-757, Republic of Korea.

Eui Jeong Hong. National Park Research Institute, Korea National Park Service, Wonju, Gangwon-do 570-811, Republic of Korea.

Corresponding author: Iksoo Kim (e-mail: ikkim81@chonnam.ac.kr).

Abstract: We sequenced the complete mitochondrial (mt) genome of *Camponotus atrox* (Hymenoptera: Formicidae), which is only distributed in Korea. The genome was 16,540 bp in size and contained typical sets of genes (13 protein-coding genes, 22 tRNAs, and 2 rRNAs). The *C. atrox* A+T-rich region, at 1,402 bp, was the longest of all sequenced ant genomes and was composed of an identical tandem repeat consisting of six 100-bp copies and one 96-bp copy. A total of 315 bp of intergenic spacer sequence was spread over 23 regions. An alignment of the spacer sequences in ants was largely feasible among congeneric species, and there was substantial sequence divergence, indicating their potential use as molecular markers for congeneric species. The A/T contents at the first and second codon positions of protein-coding genes (PCGs) were similar for ant species, including *C. atrox* (73.9% vs. 72.3%, on average). With increased taxon sampling among hymenopteran superfamilies, differences in the divergence rates (i.e., the non-synonymous substitution rates) between the suborders Symphyta and Apocrita were detected, consistent with previous results. The *C. atrox* mt genome had a unique gene arrangement, *trnI-trnM-trnQ*, at the A+T-rich region and ND2 junction (underline indicates inverted gene). This may have originated from a tandem duplication of *trnM-trnI*, resulting in *trnM-trnI-trnM-trnI-trnQ*, and the subsequent loss of the first *trnM* and second *trnI*, resulting in *trnI-trnM-trnQ*.

Key words: mitochondrial DNA, mt genome, Hymenoptera, ants, *Camponotus atrox*, gene rearrangement

¹Supplementary data are available with the article at the journal website: <http://nrcresearchpress.com>.

Introduction

Insect mitochondrial (mt) genome sequences have been compiled for more than 500 species covering all insect orders (Feijao et al. 2006). One of the most species-rich orders, Hymenoptera, is only represented by ~55 complete or nearly complete mt genomes. This includes 13 mt genomes from 11 species of ants (family Formicidae) [two sequences each from *Solenopsis invicta* (PMS and VMS) and *Pristomyrmex punctatus* (normal and cheater morphs)] in eight genera and three subfamilies, Formicinae, Myrmicinae, and Dolichoderinae (Gotzek et al. 2010; Hasegawa et al. 2011; Williams and Wernegreen 2013; Babbucci et al. 2014; Berman et al. 2014; Rodovalho et al. 2014; Tang et al. 2014; Yang et al. 2015). Ten sequences are complete and contain typical sets of genes [13 protein-coding genes (PCGs), two rRNAs, and 22 tRNAs] and a major non-coding element called the A+T-rich region found in insects. The remaining three genomes are incomplete. *Camponotus chromaiodes* lacks the sequence from *trnL_I* to *trnQ* (Williams and Wernegreen 2013) and *Polyrhachis dives* lacks the sequence spanning *lrRNA* to *trnW*, including the *ND2* gene (Tang et al. 2014). These two species belong to Formicinae; accordingly, *Formica fusca* and *F. selysi* are the only two species in Formicinae that currently have complete mt genome sequences.

Generally, insect mt genomes are arranged in a typical order that is referred to as the ancestral arrangement for insects (Boore 1999). However, Hymenoptera has exceptional and abundant gene rearrangements, similar to those of the hemipteroids (Wei et al. 2009, 2010a; Dowton et al. 2002; Cameron 2014). In fact, the *Apis mellifera* mt genome, which was the first complete hymenopteran mt genome, shows extensive gene rearrangement compared to the ancestral sequence (Crozier and Crozier 1993). At the A+T-rich region and *ND2* junction, it has *trnE-trnS_I-trnM-trnQ-trnA-trnI* instead of *trnI-trnQ-trnM* (the underline indicates an inverted gene); at the *COII* and *ATP8* junction, it has *trnD-trnK* instead of *trnK-trnD*; at the

ND3 and *ND5* junction, it has *trnR-trnN-trnF* instead of *trnA-trnR-trnN-trnS₁-trnE-trnF*. Subsequent studies have shown that the incidence of genome rearrangements is widespread in Hymenoptera (e.g., Dowton and Austin 1999; Cha et al. 2007). In particular, Dowton et al. (2009) have identified 67 gene rearrangements in 14 hymenopteran species. Nevertheless, new arrangements are continuously reported (e.g., Mao et al. 2015).

In ants, three gene arrangements have been reported to date (Babbucci et al. 2014). Species in the subfamilies Dolichoderinae and Formicinae share an identical gene arrangement, and two distinct arrangements exist in Myrmicinae (Gotzek et al. 2010; Hasegawa et al. 2011; Babbucci et al. 2014; Berman et al. 2014; Rodovalho et al. 2014; Yang et al. 2015). In other Formicinae species (e.g., *C. chromaiodes* and *P. dives*), detailed descriptions of gene arrangements are lacking due to the unavailability of complete genome sequences. Thus, it would be intriguing to explore the taxonomic extent of the gene arrangements found in *F. fusca* and *F. selysi* in other Formicinae.

In this study, we sequenced the complete mt genome of *C. atrox* in Formicinae. *Camponotus* is a large, species-rich genus, possessing more than 1,000 species, known as carpenter ants (Bolton 1995). In Korea, at least 12 *Camponotus* species, including *C. atrox*, have been identified (Kim and Kim 2006). *C. atrox* is only distributed in Korea and is often found at an altitude of 500 meters in South Korea. The complete mt genome sequence of *C. atrox* was compared to those of other ants and hymenopteran groups in terms of genome organization, arrangement, and sequence composition. Herein, we report a new gene arrangement that has not previously been detected in Hymenoptera, including Formicidae.

Materials and methods

Genomic DNA extraction

A colony of *C. atrox* was collected from Mt. Sokrisan, Korea in July, 2013 and placed in 95% alcohol. After an individual *C. atrox* was placed in a 1.5-mL microcentrifuge tube containing phosphate-buffered saline solution, the tube was inverted several times and incubated for 30 min at 55°C with the tube cap opened. Legs were removed from the body and chopped with a razor blade in a lysis buffer obtained from the Wizard™ Genomic DNA Purification Kit (Promega, Madison, WI, USA). The other DNA extraction procedures were performed following the manufacturer's instructions (Promega).

Long PCR and sequencing

One short fragment of *COI* was amplified using the primer sets reported by Folmer et al. (1994) and *CytB* primers were designed from the alignment of conspecific *C. chromaiodes* and *C. herculeanus* in this study (Savolainen and Vepsäläinen 2003). These short fragments were amplified with AccuPower® PCR PreMix (Bioneer, Daejeon, Korea) under the following conditions: initial denaturation for 4 min at 94°C, followed by 35 cycles of 30 s at 94°C, 40 s at 50–55°C, and 60 s at 72°C, with a final 7-minute extension at 72°C. On the basis of the sequence information, two pairs of primers were designed to amplify two overlapping long fragments using LA Taq™ (Takara Biomedical, Otsu, Japan) under the following conditions: initial denaturation for 2 min at 96°C, followed by 30 cycles of 10 s at 98°C and 15 min at 58–65°C, and a final 10-min extension at 72°C. The primer information is provided in Table 1.

These PCR products were then used to construct a shotgun library. In brief, DNAs were sheared into 1–5-kb fragments using Hydroshear (Gene Machine, Victoria, TX, USA), and the DNA fraction was collected using a Chromaspin TE 1000 column. The DNA fraction was then cloned into the pUC118 vector (Takara Biomedical), and the resultant plasmid DNAs were isolated with the Wizard Plus SV Minipreps DNA Purification System (Promega). DNA

sequencing was conducted with the ABI PRISM® BigDye® Terminator v3.1 Cycle Sequencing Kit and the ABI PRISM™ 3100 Genetic Analyzer (PE Applied Biosystems, Waltham, MA, USA). Two overlapping long fragments were sequenced each from 24 short fragments of ~750 bp, with 150–55 bp of overlapping sequence. The sequence of each short fragment was searched to confirm its appropriate position and sequence homology to other ants using Blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Each short fragment was then assembled into a contig, and the two contigs were assembled into the complete mt genome using SeqMan in the Lasergene software package ver. 7.1 (DNASTAR Inc., Madison, WI, USA) (Burland 2000).

Genome annotation

The *C. atrox* mt genome was generally annotated following the procedure proposed by Cameron (2014). First, the identification, boundary delimitation, and secondary structure prediction for tRNAs were performed using tRNAscan-SE 1.21 with the default search mode, Mito/Chloroplast set as the source, invertebrate mt genomes set as the genetic code for tRNA isotype prediction, and a Cove score cut-off of 1 (Lowe and Eddy 1997). Consequently, 20 tRNAs were detected. Additionally, *trnS_I*, which had a truncated DHU arm and *trnR* (71 bp), which had longer variable loop were found from hand-drawn secondary structures based on alignments with the predicted regions of *trnS_I* and *trnR* of other ant species, particularly with consideration of the anticodons (Wolstenholme 1992; Cha et al. 2007). Sequence alignments were performed using MAFFT ver. 6 (Katoh et al. 2002) with the gap opening penalty set to 1.53 and the offset value ($\approx$ gap extension penalty) set to 0.5. Second, the identification and boundary delimitation of individual mt PCGs were performed using blastx and tblastn. Start and stop codons of PCGs were further confirmed based on the alignment of *C. atrox* PCGs with those of other ants using MAFFT ver. 6 (Katoh et al. 2002). Third, two rRNAs and the

A+T-rich region were identified and delimited using Blast and were further confirmed based on the alignment with other ant mt rRNA genes and A+T-rich region sequences using MAFFT ver. 6 (Katoh et al. 2002). The sequence data were deposited in the GenBank database under the accession no. KT159775.

Comparative mt gene analyses

To compare individual gene size, location, and the sizes of overlapping and intergenic spacer sequences, as well as the presence of repeat sequences in the A+T-rich region, 13 available ant mt genomes were downloaded from GenBank or AMiGA (Feijao et al. 2006). Gene overlap and intergenic spacer sequences were manually counted. For the comparative analysis of PCGs, the taxonomic scope was extended to include Hymenoptera species. Hymenopteran superfamilies with available complete mt PCGs for more than two species were considered. The genomes of 26 species in eight superfamilies of Hymenoptera were downloaded from GenBank or AMiGA (Feijao et al. 2006). For ants, six species representing each genus were included. These 33 mt genome sequences were subjected to an analysis of the frequent codons, nucleotide composition, skewness among codon positions, and nonsynonymous (K_a) and synonymous (K_s) substitution rates. Codon usage and nucleotide composition were determined by MEGA 6 (Tamura et al. 2013). For the analyses of skewness, and K_a and K_s among hymenopteran superfamilies, Formicidae was regarded as an independent group. Thus, Vespoidea was composed of wasps, but not ants. Skewness was calculated for full-length PCGs and each codon position of the PCGs using the EditSeq program and the following formula proposed by Perna and Kocher (1995): GC-skew = $(G - C)/(G + C)$ and AT-skew = $(A - T)/(A + T)$, where C, G, A, and T represent the frequencies of the four bases. K_a and K_s were estimated to determine the degree of genetic divergence among hymenopteran superfamilies using a model that averages parameters across 14 candidate

models (Posada and Buckley, 2004) with the KaKs Calculator (v. 1.2) (Zhang et al. 2006). The *Drosophila melanogaster* mt genome (Lewis et al. 1995) was used as a criterion sequence for the calculations following Dowton et al. (2009). Subsequently, the Ka estimates among superfamilies (one species per genus from Apoidea, Ceraphronoidea, Evanioidea, Ichneumonoidea, Formicidae, Formicidae-excluded Vespoidea, Platygastroidea, Proctotrupoidea, and Tenthredinoidea) were subjected to an analysis of variance (ANOVA) implemented in JMP software ver. 11.1.1 (SAS Institute, Cary, NC, USA). The statistical significance of pairwise comparisons on composition among superfamilies was evaluated by the Tukey-Kramer HSD test.

Results and discussion

General characterization of the *Camponotus atrox* mt genome

The mt genome of *C. atrox* was 16,540 bp in size and contained typical sets of mt genes (13 PCGs, 22 tRNAs, and 2 rRNAs) and the major noncoding A+T-rich region found in insects (Table 2). The predicted secondary structures of the 22 tRNAs are presented as supplementary data (gen-2015-0080.R1 Supplement Figure 1)¹. Extra tRNAs or pseudo-tRNAs intermittently detected in other Hymenoptera were not found in the *C. atrox* mt genome (Cha et al. 2007; Wang et al. 2013). All 13 PCGs of *C. atrox* began with the typical start codon ATN and ended with TAA (Table 2). In other ants, typical ATN start codons are used. However, *Leptomyrmex pallens* *ND1* begins with TTG, which is rare in general, but canonical for invertebrate mt PCGs (supplementary data, gen-2015-0080.R1 Supplement Table 1)¹. Unlike *C. atrox*, other ant species often have infrequent TAG start codons and incomplete T or TA stop codons (supplementary data, gen-2015-0080.R1 Supplement Table 1)¹.

The sizes of most *C. atrox* mt genes were consistent with those of other ant species, but a few exceptions were also detected. *ND5* and *trnV* were the shortest in ants as 1,653 bp (range in ants, 1,656–1,671 bp) and 64 bp (66–82 bp), respectively (supplementary data, gen-2015-0080.R1 Supplement Table 1)¹. The *trnP* and *trnL_I* genes were the longest as 95 bp (range in ants, 64–88 bp) and 90 bp (55–83 bp), respectively, in ants. The shortest *trnV* formed a typical clover-leaf secondary structure with an invariable 7 bp in the aminoacyl stem, 5 bp in the anticodon stem, and 7 bp in the anticodon loop (data not shown). The longest loci *trnP* and *trnL_I* formed atypical secondary structures with 28- and 21-bp-long variable loops, respectively, and invariable aminoacyl stems, anticodon stems, and anticodon loops that were highly conserved in length (Fig. 1). A search for such unusual tRNAs in other ants resulted in *trnL_I* of *F. fusca* and *F. selysi*, which have 21- and 23-bp-long variable loops, respectively (Fig. 1). The lengths of these variable loops are extraordinary; the homologous region of other ant *trnP* and *trnL_I* ranged from only 4–6 bp and 4–8 bp, respectively (data not shown). Their functionality as tRNAs is unclear, but in yeast, the region has little functional significance based on a comparison of *trnF* with longer and shorter variable loops (Brennan and Sundaralingam 1976).

A+T-rich region

The major non-coding sequence, the A+T-rich region, was located between *srRNA* and *trnI* and was 1,402 bp in the *C. atrox* mt genome (Table 2); it was the longest of all sequenced ants (supplementary data, gen-2015-0080.R1 Supplement Table 1)¹. This region contained an identical tandem repeat consisting of six duplicated 100-bp copies (called RA1–RA6) and one 96-bp copy that lacked 4 bp at the end (called RA7); these repeats were surrounded at each end by 70- and 636-bp-long adjoining non-repeat sequences (NRS) (Fig. 2A). This repeat sequence consisted of 67% A/T nucleotides, whereas the remaining NRS consisted of 92%

A/T nucleotides, with an overall average of 79.5%. The substantial difference in A/T content between the repeat region and the remaining NRS suggests that the repeat region did not originate from the A+T-rich region. However, the A/T contents of all other regions in the *C. atrox* mt genome were also substantially different from that of the repeat region (78.8% in the whole genome, 76.3% in PCGs, 85.1% in *srRNA*, and 85.2% in *lrRNA*) (Table 3). A search of public databases using Blast and comparisons to other regions of the current *C. atrox* mt genome using MAFFT failed to identify similar sequences. Thus, the identity of the repeat region remains unknown and is an intriguing question. In addition to *C. atrox*, *F. fusca* and *F. selysi* belonging to Formicinae and *Pristomyrmex punctatus* belonging to Myrmicinae also had repeat sequences in the A+T-rich region (Fig. 2B, 2C, and 2D). Unlike the other species, *F. fusca* and *F. selysi* both had two different repeat units. The three currently available complete mt genome sequences of Formicinae (*C. atrox*, *F. fusca*, and *F. selysi*) all have repeat sequences in the A+T-rich region, but extended taxon sampling from Formicinae is required to generalize this as a feature of the subfamily.

Non-coding spacer sequences

The *C. atrox* mt genes were interleaved with 315 bp of intergenic spacer sequences spread over 23 regions, ranging in size from 1 to 66 bp. Two relatively longer regions were located between *trnM* and *trnQ* (57 bp) and *trnQ* and *ND2* (66 bp) (termed spacers 1 and 2, respectively) and the remaining spacer sequences were a maximum of ~27 bp (supplementary data, gen-2015-0080.R1 Supplement Table 2)¹. Spacer 1, which had an A/T content of 68.4%, was unique in *C. atrox* mt genome owing to the distinct gene arrangement in this region (see Gene arrangement section). Spacer 2 was the longest (66 bp) and was composed solely of A/T nucleotides. All sequenced mt genomes of ants have relatively longer spacers in this region that vary in length (50 bp in *Myrmica scabrinodis* to 189 bp in *Atta laevigata*) and G/C

content.

Intergenic spacer sequences may be difficult to align among divergent species owing to the high sequence variability and insertions/deletions. However, it is feasible to align the spacer sequences among closely related taxa. In an alignment of spacer sequences of longer than 10 bp in ants, reasonably high sequence homology was obtained for several sequences, mainly among congeneric species (supplementary data, gen-2015-0080.R1 Supplement Figure 2)¹. For example, a spacer sequence located between *COI* and *trnL₂* that is found in six ant species is longer than 10 bp (20 bp in *C. atrox*, 20 bp in *C. chromaiodes*, 130 bp in *P. dives*, 26 bp in *F. fusca*, 28 bp in *F. selysi*, and 160 bp *A. laevigata*). Only sequence alignments for the congeneric pair *C. atrox* and *C. chromaiodes*, and the congeneric *F. fusca* and *F. selysi* were possible. For species in different genera, the spacer could not align due to substantial differences in length or sequence divergence. Similarly, for six spacer sequences in *Camponotus*, 26 in *Formica*, and five in *Solenopsis*, only those of congeneric species were successfully aligned. The sequence homology of the aligned sequences ranged from 46.5% (for the spacer sequences between *trnP* and *ND6* in *Solenopsis*) to 100% (for the spacer sequences between *CytB* and *trnS₂* in *Formica*) (supplementary data, gen-2015-0080.R1 Supplement Figure 2)¹. Some of these sequences may be highly valuable as congeneric molecular markers owing to the reasonable alignment among congeneric species and substantial sequence divergence between species. The importance of spacer sequences as molecular markers for several taxonomic groups has been proposed (Fujita et al. 2004; Creer et al. 2006). Within Hymenoptera, an ~89-bp-long mt spacer sequence located between *trnL₂* and *COII* in *Apis* species is used extensively as a molecular marker for examining the biogeographic history and genetic diversity of the Asian cavity-nesting honey bee *A. cerana*, which is distributed in temperate and tropical Asia (Smith et al. 2000; Tan et al. 2007).

Nucleotide composition

The nucleotide composition of the *C. atrox* mt genome was biased toward A/T nucleotides (78.8%), similar to the genomes of other hymenopteran species, including ants (Table 3). In ants, the A/T content ranged from 69.5% (*Leptomyrmex pallens*) to 83.4% (*F. fusca*) and in ant-excluded hymenopterans, it ranged from 77.2% (*Monocellicampa pruni*) to 87.4% (*Diadegma semiclausum*) (Table 3). This estimate in ants was slightly lower than that of other hymenopterans and a similar result also was presented in a previous study (Gotzek et al. 2010). The A/T content varied substantially between RNAs (84.8% in tRNAs, 85.2% in *lrRNA*, and 85.1% in *srRNA*) and PCGs (76.3%) in the *C. atrox* mt genome; this trend was consistently observed in sequenced hymenopterans, including ants (Table 3). The biased usage of A/T nucleotides was also reflected in the codon usage analyses (Table 4). Among the 64 available codons, *C. atrox* utilized ATT (isoleucine), TTA (leucine), TTT (phenylalanine), and ATA (methionine), which are composed solely of A/T nucleotides, most frequently, accounting for 36.30% of all codons, with an average in ants of 36.57% and in ant-excluded hymenopterans of 42.24%. *L. pallens* (Dolichoderinae), which had the lowest genome-wide A/T content (69.5%), had the lowest frequency for the four codons, i.e., 27.34% (Table 4). Currently, *L. pallens* is the only species with a sequenced genome in the subfamily. Thus, it would be interesting to collect more data for this subfamily.

The nucleotide composition of the 13 concatenated PCGs in the *C. atrox* mt genome was as follows: A, 32.2%; T, 44.1%; C, 12.8%; and G, 10.9% (supplementary data, gen-2015-0080.R1 Supplement Table 3)¹. The third codon position of the PCGs (79.8%) harbored a higher A/T content than the first (75.4%) and second (73.8%) codon positions. A similar pattern was also detected in the genomes of other sequenced ants, i.e., they had an A/T content of 73.7% at the first position, 72.0% at the second position, and 83.4% at the third position, on average (supplementary data, gen-2015-0080.R1 Supplement Table 3)¹. However,

ant-excluded hymenopterans had different A/T compositions at the first and second codon positions, i.e., 77.7% and 74.0%, respectively, on average (supplementary data, gen-2015-0080.R1 Supplement Table 3)¹. This trend has also been found in the mt genomes of *A. mellifera*, *Melipona bicolor*, and *Bombus ignitus* in Apoidea (Crozier and Crozier 1993; Cha et al. 2007; Silvestre et al. 2008). The apocritan *Primeuchroeus* spp. (Chrysidoidea) had A/T contents of 72.2%, 72.1%, and 82.7% at the first, second, and third codon positions, respectively, indicating that the first and second positions were nearly identical (Castro et al. 2006). As a whole, ants have similar A/T contents at the first and second codon positions, and other hymenopterans have different A/T contents at these positions, although exceptions exist.

Compositional skew

The nucleotide bias was also reflected in the form of base skewness (supplementary data, gen-2015-0080.R1 Supplement Table 4)¹. The *C. atrox* mt genome was moderately T-skewed (-0.156) and slightly C-skewed (-0.079). Considering all codon positions, the nucleotides were T-skewed to varying degrees (-0.003 at first, -0.396 at second, and -0.077 at third codon positions), whereas nucleotides at first codon positions were G-skewed (0.075) and those at second and third codon positions were C-skewed (-0.135 and -0.194, respectively) in the *C. atrox* mt genome. A similar skewness pattern was observed among ants, but *C. atrox* and *F. fusca* in Formicinae were T-skewed at the first codon position (-0.003 and -0.013, respectively), whereas other ants were slightly A-skewed at the first position. The strong T-skewness at the second codon position (-0.389 to -0.439) in ants, including *C. atrox*, is noteworthy because it differed completely from the first codon position, which showed a slight A or T skewness (-0.013 to 0.057), despite their nearly equivalent non-degenerative status compared to the third codon position and the similar A/T content at the two positions in ants. The evolutionary pressure that favors Ts is more severe at the second codon position

than the first, although such codon position-based inequalities have yet to be clearly understood.

Representative hymenopterans, including ants, showed a similar pattern with respect to codon position-based skewness, with some obvious exceptions. When the skew values were plotted in an ellipse that covers 90% of A/T and G/C skewness within the normal density, a few species deviated from this coverage (Fig. 3). These included *Cotesia vestalis* and *Spathius agrili* in Ichneumonoidea and *Gasteruption* sp. in Evanioidea for the concatenated PCGs; *S. agrili* and *C. vestalis* in Ichneumonoidea and *Wallacidia oculata* in Vespoidea for the first codon position; *L. pallens* in Formicidae and *Vanhornia eucnemidarum* in Proctotrupoidea for the second codon position; and *Colletes gigas* in Apoidea, *C. vestalis* in Ichneumonoidea, and *V. eucnemidarum* in Proctotrupoidea for the third codon position. Currently, data regarding the mt genomes at the family or subfamily levels are extremely limited for these species. Thus, further detailed examinations of a wider range of taxa are required.

Genetic divergence

Ka and Ks were estimated to infer the extent of genetic divergence among species. Ka ranged from 0.25463 (*Allantus luctifer* in Tenthredinoidea) to 0.56365 (*W. oculata* in Vespoidea); Ks ranged from 4.57691 (*L. pallens* in Formicidae) to 9.14568 (*Trissolcus basalis* in Platygastroidea); Ka/Ks ranged from 0.03066 (*A. luctifer* in Tenthredinoidea) to 0.11260 (*W. oculata* in Vespoidea) (supplementary data, gen-2015-0080.R1 Supplement Table 5)¹. These comparisons indicate that Ks was saturated. The two earliest diverging symphytan species *A. luctifer* and *Monocellicampa pruni* in Tenthredinoidea had the lowest Ks values of 0.25463 and 0.27200, respectively, while other bees belonging to the suborder Apocrita tended to have higher Ks values than those of members of the suborder Symphyta (0.36649–0.56365). Previously, different rates of divergence between Apocrita and Symphyta

have been reported and these previous findings are supported by the current analysis, which includes more extensive taxon sampling (from 14 to 33 species) (Dowton and Austin 1995; Dowton et al. 2009). Within Formicidae, K_a ranged from 0.42518 (*M. scabrinodis*) to 0.45342 (*P. punctatus*), indicating similar divergence levels among ant species (supplementary data, gen-2015-0080.R1 Supplement Table 5)¹. To confirm the differences in the rate of genetic divergence, K_a was compared at the superfamily level using ANOVA (Fig. 4). A pairwise comparison of K_a among hymenopteran superfamilies confirmed that the early diverging symphytan Tenthredinoidea species differed significantly from other hymenopteran superfamilies. This result is consistent with the previous findings that different rates of divergence exist between Apocrita and Symphyta (Dowton and Austin 1995; Dowton et al. 2009). Excluding the symphytan Tenthredinoidea species, two pairs of superfamilies differed significantly in K_a , i.e., Vespoidea and Ichneumonoidea ($p = 0.0102$) and Apoidea and Ichneumonoidea ($p = 0.0054$). These results indicate that members of Symphyta are not the only group with statistically different degrees of divergence in Hymenoptera.

Gene arrangement

The orientation and gene order of the *C. atrox* mt genome differed from those of other hymenopterans, including ants; their genome contained *trnI-trnM-trnQ* in the A+T-rich region and ND2 junction (Fig. 5; Dowton and Austin 1999; Wei et al. 2009). To date, three different mt genome arrangements have been reported in ants. Two Formicinae species (*F. fusca* and *F. selysi*) and *L. pallens* in Dolichoderinae have *trnM-trnI-trnQ* in the A+T-rich region and ND2 junction (Babbucci et al. 2014; Berman et al. 2014; Yang et al. 2015). All sequenced Myrmicinae (*Solenopsis richteri*, *S. geminate*, *S. invicta*, *A. laevigata*, *M. scabrinodis*, and *P. punctatus*) have *trnV-trnM-trnI-trnQ* in the A+T-rich region and ND2 junction owing to a transposition of *trnV* from the lrRNA and srRNA junction (Babbucci et al.

2014; Gotzek et al. 2010; Hasegawa et al. 2011; Rodovalho et al. 2014). *P. punctatus* in Myrmicinae has an additional rearrangement at the COII and ATP8 junction, i.e., *trnD-trnK*, instead of *trnK-trnD*. Because mt genome sequences of ants are only available for 3 of ~20 subfamilies, it is not possible to accurately infer the ancestral arrangement. Nevertheless, when the arrangements found in ants were mapped on the phylogenetic tree presented in Ward (2007), *trnM-trnI-trnQ* in the A+T-rich region and ND2 junction is plesiomorphic, at least for the sequenced ant subfamilies; it was found in Formicinae and Dolichoderinae. If this is the most parsimonious explanation, the current *C. atrox* arrangement can be explained by a tandem duplication of *trnM-trnI*, resulting in *trnM-trnI-trnM-trnI* and a loss of the first *trnM* and the second *trnI*, ultimately resulting in *trnI-trnM-trnQ* (Fig. 6). Similarly, the *trnV-trnM-trnI-trnQ* found exclusively in Myrmicinae species can be explained by a tandem duplication of the *trnV-srRNA*-A+T-rich region, resulting in a *trnV-srRNA*-A+T-rich region-*trnV-srRNA*-A+T-rich region, and a loss of the first *trnV* and second *srRNA*-A+T-rich region, resulting in *srRNA*-A+T-rich region-*trnV-trnM-trnI-trnQ*.

At the COII and ATP8 junction, Dowton and Austin (1999) have determined that the *trnK-trnD* arrangement is ancestral, although both ancestral and derived arrangements (*trnD-trnK*) are found in various monophyletic hymenopteran groups, indicating that the rearrangements are independently derived. In ants, the derived arrangement is found only in a species of Myrmicinae (*P. punctatus*), whereas the ancestral *trnD-trnK* arrangement is found in all other ant species, including other species of Myrmicinae (Fig. 6). A tandem duplication of *trnK-trnD*, resulting in *trnK-trnD-trnK-trnD* and a loss of the first *trnK* and the second *trnD* appear to explain current *trnD-trnK* in *P. punctatus*. As additional mt genome sequences from ants become available, advanced analyses of the evolutionary history of gene rearrangements will be possible.

Acknowledgments

This research was supported by a research grant from the National Institute of Biological Resources, “Origin of Biological Diversity of Korea: Molecular Phylogenetic Analyses of Major Korean Taxa,” awarded to Iksoo Kim.

References

- Babbucci, M., Basso, A., Scupola, A., Patarnello, T., and Negrisol, E. 2014. Is it an ant or a butterfly? Convergent evolution in the mitochondrial gene order of Hymenoptera and Lepidoptera. *Genome Biol. Evol.* **6(12)**: 3326-3343. doi:10.1093/gbe/evu265.
- Berman, M., Austin, C.M., and Miller, A.D. 2014. Characterisation of the complete mitochondrial genome and 13 microsatellite loci through next-generation sequencing for the new caledonian spider-ant *Leptomyrmex pallens*. *Mol. Biol. Rep.* **41(3)**: 1179-1187. 10.1007/s11033-013-2657-5
- Bolton, B. 1995. A new and general catalogue of the ants of the world. Harvard Univ. Press, Cambridge, M.A.
- Boore, J.L. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* **27(8)**: 1767-1780. doi:10.1093/nar/27.8.1767.
- Burland, T.G. 2000. DNASTAR's laser gene sequence analysis software. In *Bioinformatics methods and protocols*. Edited by S. Misener, S.A. Krawets. Humana Press, Inc., Totowa, NJ. Pp. 71-91.
- Brennan, T., and Sundaralingam, M. 1976. Structure of transfer RNA molecules containing the long variable loop. *Nucleic Acid Res.* **3(11)**: 3235-3252. doi:10.1093/nar/3.11.3235.
- Cameron, S.L. 2014. Insect mitochondrial genomics: implications for evolution and

- phylogeny. *Annu. Rev. Entomol.* **59**: 95-117. doi:10.1146/annurev-ento-011613-162007.
- Cameron, S.L., Dowton, M., Castro, L.R., Ruberu, K., Whiting, M.F., Austin, A.D., and Stevens, J. 2008. Mitochondrial genome organization and phylogeny of two vespid wasps. *Genome* **51(10)**: 800-808. doi:10.1139/G08-066.
- Castro, L.R., Ruberu, K., and Dowton, M. 2006. Mitochondrial genomes of *Vanhornia eucnemidarum* (Apocrita: vanhorniidae) and *Primeuchroeus* spp. (Aculeata: chrysididae): evidence of rearranged mitochondrial genomes within the Apocrita (Insecta: hymenoptera). *Genome* **49(7)**: 752-766. doi:10.1139/g06-030.
- Cha, S.Y., Yoon, H.J., Lee, E.M., Yoon, M.H., Hwang, J.S. Jin, B.R., Han, Y.S., and Kim I. 2007. The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae). *Gene* **392(1-2)**: 206-220. doi:10.1016/j.gene.2006.12.031.
- Creer, S., Pook, C.E., Malhotra, A., and Thorpe, R.S. 2006. Optimal intron analyses in the *Trimeresurus* radiation of Asian pitvipers. *Syst. Biol.* **55(1)**: 57-72. doi:10.1080/10635150500431213.
- Crozier, R.H., and Crozier, Y.C. 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* **133(1)**: 97-117.
- Dowton, M., and Austin, A.D. 1995. Increased genetic diversity in mitochondrial genes is correlated with the evolution of parasitism in the Hymenoptera. *J. Mol. Evol.* **41(6)**: 958-965. doi:10.1007/BF00173176.
- Dowton, M., and Austin, A.D. 1999. Evolutionary dynamics of a mitochondrial rearrangement 'hotspot' in the Hymenoptera. *Mol. Biol. Evol.* **16(2)**: 298-309.
- Dowton, M., Cameron, S.L., Dowavic, J.I., Austin, A.D., and Whiting, M.F. 2009. Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. *Mol. Biol. Evol.*

26(7): 1607-1617. doi: 10.1093/molbev/msp072.

Dowton, M., Castro, L.R., and Austin, A.D. 2002. Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of genome 'morphology'.

Invertebr. Syst. **16(3)**: 345-356. doi:10.1071/IS02003.

Feijao, P.C., Neiva, L.S., Azeredo-Espin, A.M.L., and Lessinger, A.C. 2006. AMiGA: the arthropodan mitochondrial genomes accessible database. Bioinformatics **22(7)**: 902-903.

doi:10.1093/bioinformatics/btl021. Available from <http://amiga.cbmeg.unicamp.br/> [accessed 14 January 2005].

Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. **3(5)**: 294-299.

Fujita, M.K., Engstrom, T.N., Starkey, D.E., and Shaffer, H.B. 2004. Turtle phylogeny: insights from a novel nuclear intron. Mol. Phylogenet. Evol. **31(3)**: 1031-1040. doi:10.1016/j.ympev.2003.09.016.

Gibson, A., Gowri-Shankar, V., Higgs, P.G., and Rattray, M. 2005. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. Mol. Biol. Evol. **22(2)**: 251-264. doi:10.1093/molbev/msi012.

Gissi, C., Reyes, A., Pesole, G., and Saccone, C. 2000. Lineage-specific evolutionary rate in mammalian mtDNA. Mol. Biol. Evol. **17(7)**: 1022-1031.

Gotzek, D., Clarke, J., and Shoemaker, D. 2010. Mitochondrial genome evolution in fire ants (Hymenoptera: Formicidae). BMC Evol. Biol. **10**: 300. doi:10.1186/1471-2148-10-300.

Hasegawa, E., Kobayashi, K., Yagi, N., and Tsuji, K. 2011. Complete mitochondrial genomes of normal and cheater morphs in the parthenogenetic ant *Pristomyrmex punctatus* (Hymenoptera: Formicidae). Myrmecol. News. **15**: 85-90.

- Huang, D., Su, T., Qu, L., Wu, Y., Gu, P., He, B., and Zhu, C. 2014. The complete mitochondrial genome of the *Colletes gigas* (Hymenoptera: Colletidae: Colletinae) [online]. Mitochondrial DNA, **Early Online**: 1-2. doi:10.3109/19401736.2014.987243.
- Kaltenpoth, M., Corneli, P.S., Dunn, D.M., Weiss, R.B., Strohm, E., and Seger, J. 2012. Accelerated evolution of mitochondrial but not nuclear genomes of Hymenoptera: new evidence from crabronid wasps. PLoS ONE **7(3)**: doi:e32826. 10.1371/journal.pone.0032826.
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. **30(14)**: 3059-3066. doi:10.1093/nar/gkf436.
- Kim, M.J., Kang, A.R., Jeong, H.C., Kim, K.-G., and Kim, I. 2011. Reconstructing intraordinal relationships in Lepidoptera using mitochondrial genome data with the description of two newly sequenced lycaenids, *Spindasis takanonis* and *Protantigius superans* (Lepidoptera: Lycaenidae). Mol. Phylogenet. Evol. **61(2)**: 436-445. doi:10.1016/j.ympev.2011.07.013.
- Kim, K.-G., and Kim, B.-J. 2006. Molecular phylogeny of *Camponotus* ants in Korea. Entomol. Res. **36(1)**: 48-55. doi:10.1111/j.1748-5967.2006.00009.x.
- Kwon, T.-S., Kim, S.-S., Lee, C.M., Jung, S.J., and Sung, J.H. 2012. Korean ant atlas (2006–2009). Korea Forest Research Institute. Research note. No. 459.
- Lewis, D.L., Farr, C.L., and Kaguni, L.S. 1995. *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequences and evolutionary comparisons. Insect Mol. Biol. **4(4)**: 263-278. doi:10.1111/j.1365-2583.1995.tb00032.x.
- Li, Q., Wei, S.J., Shi, M., and Chen, X.X. 2015. The mitochondrial genome of *Diadromus collaris* (Hymenoptera: Ichneumonidae). Mitochondrial DNA, **26(2)**: 303-304. doi: 10.3109/19401736.2013.825786. doi: 10.1093/nar/25.5.0955.

- Lowe, T.M., and Eddy, S.R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25(5)**: 955-964. doi:10.1093/nar/25.5.0955.
- Mao, M., Austin, A.D., Johnson, N.F., and Dowton, M. 2014a. Coexistence of minicircular and a highly rearranged mtDNA molecule suggests that recombination shapes mitochondrial genome organization. *Mol Biol Evol.* **31(3)**: 636-644. doi:10.1093/molbev/mst255.
- Mao, M., and Dowton, M. 2014. Complete mitochondrial genomes of *Ceratobaeus* sp. and *Idris* sp. (Hymenoptera: Scelionidae): shared gene rearrangements as potential phylogenetic markers at the tribal level. *Mol Biol Rep.* **41(10)**: 6419-6427. doi:10.1007/s11033-014-3522-x.
- Mao, M., Gibson, T., and Dowton, M. 2014b. Evolutionary dynamics of the mitochondrial genome in the Evaniomorpha (Hymenoptera)—a group with an intermediate rate of gene rearrangement. *Genome Biol Evol.* **6(7)**: 1862-1874. doi:10.1093/gbe/evu145. .
- Mao, M., Gibson, T., and Dowton, M. 2015. Higher-level phylogeny of the Hymenoptera inferred from mitochondrial genomes. *Mol. Phylogenet. Evol.* **84**: 34-43. doi:10.1016/j.ympev.2014.12.009.
- Mao, M., Valerio, A., Austin, A.D., Dowton, and M., Johnson, N.F. 2012. The first mitochondrial genome for the wasp superfamily Platygastroidea: the egg parasitoid *Trissolcus basalis*. *Genome* **55(3)**: 194-204. doi:10.1139/g2012-005.
- Perna, N.T., and Kocher, T.D. 1995. Patterns of nucleotide composition of four fold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* **41(3)**: 353-358. doi:10.1007/BF00186547.
- Rodvalho, C. M., Lyra M.L., Ferro, M., and Bacci, M. Jr. 2014. The mitochondrial genome of the leaf-cutter ant *Atta laevigata*: a mitogenome with a large number of intergenic

- spacers. PLoS One **9**(5): e97117. doi:10.1371/journal.pone.0097117.
- Savolainen, R., and Vepsäläinen, K. 2003. Sympatric speciation through intraspecific social parasitism. Proc. Natl. Acad. Sci. U.S.A. **100**(12): 7169-7174. doi: 10.1073/pnas.1036825100.
- Silverio, M.S., Rodovalho, V.D.R., Bonetti, A.M., de Oliveira, G.C., Cuadros-Orellana, S., Ueira-Vieira, C., and Rodrigues dos Santos, A. 2014. Preliminary characterization of mitochondrial genome of *Melipona scutellaris*, a Brazilian stingless bee. Biomed Res. Int. **2014**: 927546. doi:10.1155/2014/927546.
- Silvestre, D., Dowton, M., and Arias, M.C. 2008. The mitochondrial genome of the stingless bee *Melipona bicolor* (Hymenoptera, Apidae, Meliponini): sequence, gene organization and a unique tRNA translocation event conserved across the tribe Meliponini. Gen. Mol. Biol. **31**(2): 451-460. doi:10.1590/S1415-47572008000300010.
- Smith, D.R., Villafuerte, L., Otis, G., and Palmer, M.R. 2000. Biogeography of *Apis cerana* F. and *A. nigrocincta* Smith: Insights from mtDNA studies. Apidologie **31**(2): 265-279. Doi: 10.1051/apido:2000121.
- Tan, K., Warrit, N., Smith, D.R. 2007. Mitochondrial DNA diversity of Chinese *Apis cerana*. Apidologie **38**(3): 238-246. 10.1051/apido:2007008.
- Tan, M., Zhang, R., Hardman, C., and Zhou, X. 2015. Mitochondrial genome of *Hylaeus dilatatus* (Hymenoptera: Colletidae) [online]. Mitochondrial DNA, **Early Online**: 1-2. Doi: 10.3109/19401736.2014.989511.
- Tang, M., Tan, M., Meng, G., Yang, S., Su, X., Liu, S., Song, W., Li, Y., Wu, Q., Zhang, A., and Zhou, X. 2014. Multiplex sequencing of pooled mitochondrial genomes - A crucial step toward biodiversity analysis using mito-metagenomics. Nucl. Acids Res. **42**(22): e166. doi:10.1093/nar/gku917.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. 2013. MEGA6: Molecular

- Evolutionary Genetics Analysis ver. 6.0. Mol. Biol. Evol. **30(12)**: 2725-2729. doi:10.1093/molbev/mst197.
- Wan, X., Kim, M.J., and Kim, I. 2013. Description of new mitochondrial genomes (*Spodoptera litura*, Noctuoidea and *Cnaphalocrocis medinalis*, Pyraloidea) and phylogenetic reconstruction of Lepidoptera with the comment on optimization schemes. Mol. Biol. Rep. **40(11)**: 6333-6349. doi:10.1007/s11033-013-2748-3.
- Wang, A.R., Kim, M.J., Park, J.S., Choi, Y.S., Thapam R., Lee, K.Y., and Kim I. 2013. Complete mitochondrial genome of the dwarf honeybee, *Apis florea* (Hymenoptera: Apidae). Mitochondrial DNA **24(3)**: 208-210. doi:10.3109/19401736.2012.744986.
- Wei, S.J., Li, Q., van Achterberg, K., and Chen, X.X. 2014a. Two mitochondrial genomes from the families Bethyridae and Mutillidae: Independent rearrangement of protein-coding genes and higher-level phylogeny of the Hymenoptera. Mol. Phylogen. Evol. **77**: 1-10. doi:10.1016/j.ympev.2014.03.023.
- Wei, S.J., Niu, F.F., and Du, B.Z. 2014b. Rearrangement of *trnQ-trnM* in the mitochondrial genome of *Allantus luctifer* (Smith) (Hymenoptera: Tenthredinidae) [online]. Mitochondrial DNA, **Early Online**: 1-2. doi:10.3109/19401736.2014.919475.
- Wei, S.J., Niu, F.F., and Tan, J.L. 2014c. The mitochondrial genome of the *Vespa bicolor* Fabricius (Hymenoptera: Vespidae: Vespinae) [online]. Mitochondrial DNA, **Early Online**: 1-2. doi:10.3109/19401736.2014.919484.
- Wei, S.J., Shi, M., He, J.H., Sharkey, M., and Chen, X.X. 2009. The complete mitochondrial genome of *Diadegma semiclausum* (Hymenoptera: Ichneumonidae) indicates extensive independent evolutionary events. Genome **52(4)**: 308-319. doi:10.1139/G09-008.
- Wei, S.J., Shi, M., Sharkey, M.J., van Achterberg, C., and Chen, X.X. 2010a. Comparative mitogenomics of Braconidae (Insecta: Hymenoptera) and the phylogenetic utility of mitochondrial genomes with special reference to Holometabolous insects. BMC

- genomics*, **11**: 371. doi:10.1186/1471-2164-11-371.
- Wei, S.J., Tang, P., Zheng, L.H., Shim M., and Chen, X.X. 2010b. The complete mitochondrial genome of *Evania appendigaster* (Hymenoptera: Evaniidae) has low A+T content and a long intergenic spacer between *atp8* and *atp6*. *Mol. Biol. Rep.* **37**(4): 1931-1942. doi:10.1007/s11033-009-9640-1.
- Wei, S.J., Wu, Q.L., van Achterberg, K., and Chen, X.X. 2015a. Rearrangement of the *nadl* gene in *Pristaulacus compressus* (Spinola) (Hymenoptera: Evanioidea: Aulacidae) mitochondrial genome. *Mitochondria DNA* **26**(4): 1940-1744. doi: 10.3109/19401736.2013.834436.
- Wei, S.J., Wu, Q.L., and Liu, W. 2015b. Sequencing and characterization of the *Monocellicampa pruni* (Hymenoptera: Tenthredinidae) mitochondrial genome. *Mitochondrial DNA*, **26**(1): 157-158. doi:10.3109/19401736.2013.819501.
- Ward, P.S. 2007. Phylogeny, classification and species-level taxonomy of ants (Hymenoptera: Formicidae). *Zootaxa* **1668**: 549-563.
- Williams, L.E., and Wernegreen, J.J. 2013. Sequence context of indel mutations and their effect on protein evolution in a bacterial endosymbiont. *Genome Biol. Evol.* **5**(3): 599-605. doi: 10.1093/gbe/evt033.
- Wolstenholme, D.R. 1992. Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cyt.* **141**: 173-216.
- Yang, S., Li, X., Cai, L.-G., and Qian, Z.-Q. 2015. Characterization of the complete mitochondrial genome of *Formica selysi* (Insecta: Hymenoptera: Formicidae: Formicinae) [online]. *Mitochondrial DNA*, **Early online**: 1-2. doi:10.3109/19401736.2015.1018229.
- Zhang, Z., Li, J., Zhao, X.-Q., Wang, J., Wong, G.K.-S., and Yu, J. 2006. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genom. Proteom. Bioinf.* **4**(4): 259-263. doi:10.1016/S1672-0229(07)60007-2.

Figure legends Kim et al.

Fig. 1. Atypical secondary clover-leaf structures of the tRNAs found in ants, including those of *Camponotus atrox*. Dashes (-) indicate Watson-Crick base-pairing, and centered asterisks (*) indicate G-U base-pairing. Arms of tRNAs (clockwise from top) are the aminoacyl acceptor arm, TΨC arm, anticodon arm, and DHU arm. tRNA abbreviations follow the IU-PAC-IUB one-letter code. Gene names that are not underlined indicate clockwise transcription, whereas those that are underlined are transcribed in a counter-clockwise direction.

Fig. 2. Schematic illustration of the A+T-rich region. (A) *Camponotus atrox*, (B) *Formica fusca*, (C) *F. selysi*, and (D) *Pristomyrmex punctatus*. Gene names that are not underlined are transcribed in a clockwise direction, whereas those that are underlined are transcribed in a counter-clockwise direction. tRNAs are denoted by one-letter symbols in accordance with the IUPAC-IUB single-letter amino acid codes. NRS, non-repeat sequences. Asterisks indicate identical sequences in the alignment. Bar (-) indicates deletion with regard to other repeats in

the alignment.

Fig. 3. Scatter plot of A/T and G/C skews for hymenopteran superfamilies, including ants. Ellipse indicates a bivariate normal density for 90% coverage. (A) 13 concatenated protein-coding genes (PCGs), (B) first codon position of PCGs, (C) second codon position of PCGs, and (D) third codon position of PCGs. The species that deviated from the 90% coverage area are as follows: *Cotesia vestalis* and *Spathius agrili* in Ichneumonoidea and *Gasteruption* sp. in Evanioidea for the concatenated PCGs; *S. agrili* and *C. vestalis* in Ichneumonoidea and *Wallacidia oculata* in Vespoidea for the first codon position; *L. pallens* in Formicidae and *Vanhornia eucnemidarum* in Proctotrupoidea for the second codon position; and *Colletes gigas* in Apoidea, *C. vestalis* in Ichneumonoidea, and *V. eucnemidarum* in Proctotrupoidea for the third codon position.

Fig. 4. Plot of the Ka estimates according to superfamilies. The Ka score for Tenthredinoidea species, which belong to the suborder Symphata, was significantly different from all other superfamilies in Apocrita ($p < 0.001$). Each ant-excluded Vespoidea and Apoidea differed significantly from Ichneumonoidea at $p = 0.0102$ and 0.0054 , respectively. Vertical bars indicate standard deviations.

Fig. 5. Schematic illustration of the mitochondrial gene rearrangements in Formicidae, including that of *Camponotus atrox*. Gene names that are not underlined indicate a forward transcriptional direction, whereas those that are underlined indicate a reverse transcriptional direction. Gene sizes are not drawn to scale. Only genes rearranged with respect to the ancestral arrangement are presented. Thick dotted lines indicate omitted genes between *trnQ* and *trnK* (or *trnD*). ^a*C. chromaiodes* sequence is not available from *trnL₁* to *trnQ*. ^b*Polyrhachis dives* sequence is not available from *lrRNA* to *trnW*. ^c*Atta laevigata* sequence is not available for the A+T-rich region.

Fig. 6. Mitochondrial gene arrangement for Formicidae, mapped on a phylogenetic tree from Ward (2007). †, Extinct taxa. Taxonomic names written in grey do not have available mitochondrial genomes. Inner boxes explain the tandem duplication and random loss scenarios that generated the current gene arrangements found in ants.

Table 1. List of primers used to amplify and sequence the mitochondrial genome of *Camponotus atrox*.

Fragment name	Primer name	Direction	Location ^a	Mismatch	Sequence (5'→3')	Reference
Short fragments						
SF1 (<i>COI</i>)	LCO1490	F	1580-1604	3	GGTCAACAAATCATAAAGATATTGG	Folmer et al. (1994)
"	HCO2198	R	2263-2288	1	TAAACTTCAGGGTGACCAAAAAATCA	"
SF2 (<i>CytB</i>)	FORMI-CYTB-F1	F	10920-10940	1	ATTTAAATTAATCAATACATG	Designed in this study
"	FORMI-CYTB-R1	R	11479-11501	1	ATAATAATTGATGATACAAGTGC	"
Long fragments						
LF1 (<i>COI- CytB</i>)	CAMPO-LF-COI-F2	F	1722-1742	0	TACCATCGTTACAAGTCATGC	"
"	CAMPO-LF-CYTB-R1	R	11177-11199	0	GAAATATAGGTGAATGATGACTA	"
LF1 (<i>CytB-COI</i>)	CAMPO-LF-CYTB-F3	F	11036-11058	0	TAATCACCAACCTTGTATCTACG	"
"	CAMPO-LF-COI -R4	R	2477-2497	0	TAGCGATGATTATAGTAGCTG	"

Note: ^aLocations are with respect to *C. atrox* mitochondrial genome.

Table 2. Summary of *Camponotus atrox* mitochondrial genome.

Gene	Direction	Nucleotide number	Size	Anticodon	Start codon	Stop codon
<i>trnI</i>	F	1-70	70	GAT 30-32		
<i>trnM</i>	F	82-151	70	CAT 114-116		
<i>trnQ</i>	R	209-279	71	TTG 244-246		
<i>ND2</i>	F	346-1344	999		ATT	TAA
<i>trnW</i>	F	1348-1420	73	TCA1378-1380		
<i>trnC</i>	R	1420-1485	66	GCA 1451-1453		
<i>trnY</i>	R	1488-1553	66	GTA 1521-1523		
<i>COI</i>	F	1564-3093	1530		ATG	TAA
<i>trnL₂</i>	F	3114-3182	69	TAA 3143-3145		
<i>COII</i>	F	3183-3869	687		ATT	TAA
<i>trnK</i>	F	3882-3951	70	TTT 3912-3914		
<i>trnD</i>	F	3952-4020	69	GTC 3982-3984		
<i>ATP8</i>	F	4021-4176	156		ATA	TAA
<i>ATP6</i>	F	4173-4838	666		ATA	TAA
<i>COIII</i>	F	4866-5657	792		ATG	TAA
<i>trnG</i>	F	5675-5741	67	TCC 5705-5707		
<i>ND3</i>	F	5739-6092	354		ATA	TAA
<i>trnA</i>	F	6103-6167	65	TGC 6133-6135		
<i>trnR</i>	F	6168-6238	71	TCG 6197-6199		
<i>trnN</i>	F	6228-6295	68	GTT 6259-6261		
<i>trnS₁</i>	F	6303-6360	58	TCT 6323-6325		
<i>trnE</i>	F	6363-6430	68	TTC6394-6396		
<i>trnF</i>	R	6439-6504	66	GAA 6469-6471		
<i>ND5</i>	R	6507-8159	1653		ATA	TAA
<i>trnH</i>	R	8171-8241	71	GTG 8208-8210		
<i>ND4</i>	R	8246-9586	1341		ATT	TAA
<i>ND4L</i>	R	9586-9882	297		ATA	TAA
<i>trnT</i>	F	9884-9948	65	TGT 9914-9916		
<i>trnP</i>	R	9960-10054	95	TGG 9899-9901		
<i>ND6</i>	F	10067-10615	549		ATG	TAA
<i>CytB</i>	F	10621-11736	1116		ATG	TAA
<i>trnS₂</i>	F	11745-11813	69	TGA 11775-11777		
<i>ND1</i>	R	11823-12782	960		ATA	TAA
<i>trnL₁</i>	R	12780-12869	90	TAG 12838-12840		
<i>lrRNA</i>	R	12870-14297	1428			
<i>trnV</i>	R	14298-14361	64	TAC 14329-14331		
<i>srRNA</i>	R	14362-15138	778			
A+T-rich region		15139-16540	1402			

Note: The abbreviations for the genes are as follows: *COI*, *COII* and *COIII* refer to the cytochrome oxidase subunits, *CytB* refers to cytochrome b, and *ND1-6* refers to NADH dehydrogenase components. *srRNA* and *lrRNA* denote small and large subunit ribosomal RNA (rRNA) genes. tRNAs are denoted as one-letter symbols in accordance with the IUPAC-IUB single-letter amino acid codes except those encoding leucine and serine, which are labeled *L₁* for CTN codon family, *L₂* for TTR codon family, *S₁* for AGN codon family, and *S₂* for TCN codon family.

Table 3. Characteristics of the hymenopteran mitochondrial genomes.

Taxon	Size (bp)	A/T content (%)	PCG ^b		<i>srRNA</i>		<i>lrRNA</i>		tRNA		A+T-rich region		GenBank accession no.	References
			No. codons ^a	AT%	Size (bp)	AT%	Size (bp)	AT%	Size (bp)	AT%	Size (bp)	AT %		
Hymenoptera														
Apocrita														
Aculeata														
Vespoidea														
Formicidae														
Formicinae														
Camponotini														
<i>Camponotus atrox</i>	16,540	78.8	3,664	76.3	777	85.1	1,428	85.2	1,464	84.8	1,402	79.5	KT159775	This study
<i>Camponotus chromaiodes</i> *	12,473	78.1	3,679	77.0	-	-	-	-	1,165	85.5	-	-	JX966368	Williams and Wernegreen, 2013
<i>Polyrhachis dives</i> *	13,129	78.8	3,256	77.7	-	-	203	72.4	1,154	84.6	-	-	KM244657	Tang et al. 2014
Formicini														
<i>Formica fusca</i>	16,673	83.4	3,709	81.1	802	87.8	1,502	85.6	1,433	86.6	529	86.4	LN607805	Babbucci et al. 2014
<i>Formica selysi</i>	16,752	83.3	3,707	81.0	805	87.2	1,502	86.5	1,561	86.7	585	83.4	KP670862	Yang et al. 2015
Myrmicinae														
Solenopsidini														
<i>Solenopsis richteri</i>	15,560	77.2	3682	74.9	784	83.7	1,333	80.9	1,507	85.3	379	91.8	HQ215539	Gotzek et al. 2010
<i>Solenopsis geminata</i>	15,552	76.5	3,673	73.5	783	83.3	1,338	79.9	1,449	86.1	376	92.3	HQ215537	Gotzek et al. 2010
<i>Solenopsis invicta</i> (PMS)	15,549	77.2	3,672	74.2	781	84.0	1,333	81.0	1,507	85.3	377	93.1	HQ215538	Gotzek et al. 2010
<i>Solenopsis invicta</i> ? (VMS)	15,548	77.5	3,673	74.5	784	83.3	1,332	81.5	1,509	85.0	506	91.9	HQ215540	Gotzek et al. 2010
Attini														
<i>Atta laevigata</i> *	18,729	80.8	3,692	77.6	795	85.5	1,426	83.0	1,528	86.0	-	-	KC346251	Rodvalho et al. 2014
Myrmicini														
<i>Pristomyrmex punctatus</i> (normal)	16,269	79.6	3,679	77.9	800	86.1	1,330	84.0	1,490	84.4	966	74.7	AB556946	Hasegawa et al. 2011
<i>Pristomyrmex punctatus</i> (cheaters)	16,180	79.7	3,679	77.8	800	86.3	1,327	84.0	1,490	84.4	881	75.8	AB556947	Hasegawa et al. 2011
<i>Myrmica scabrinodis</i>	15,310	82.2	3,664	80.1	784	87.4	1,296	84.0	1,481	85.8	472	96.6	LN607806	Babbucci et al. 2014
Dolichoderinae														
Dolichoderini														
<i>Leptomyrmex pallens</i>	15,591	69.5	3,691	67.3	774	64.7	1,274	74.1	1,488	79.6	195	95.4	KC160533	Berman et al. 2014
Vespidae														
Vespinae														
<i>Vespa bicolor</i>	16,937	81.7	3,731	79.3	1,451	84.8	838	85.2	1,372	85.6	898	88.2	KJ735511	Wei et al. 2014a
Eumeninae														
<i>Abispa ephippium</i>	16,953	80.6	3,756	78.6	1,368	82.9	812	80.4	1,787	83.5	308	89.9	EU302588	Cameron et al. 2008
Mutillidae														

Mutillinae															
<i>Wallacidia oculata</i>	18,442	77.4	3,537	75.0	1,247	77.4	811	74.1	1,431	80.9	3,840	83.6	FJ611801	Wei et al. 2014b	
Apoidea															
Apidae															
Apinae															
<i>Apis florea</i>	15,993	86.1	3,667	83.8	1,332	85.3	791	85.8	1,621	87.9	987	93.7	JX982136	Wang et al. 2013	
Bombinae															
<i>Bombus ignitus</i>	16,434	86.8	3,644	85.1	1,372	85.5	747	86.6	1,483	88.3	859	96.0	DQ870926	Cha et al. 2007	
Meliponinae															
<i>Melipona scutellaris</i>	14,862	86.3	3,673	86.3	1,359	86.2	756	86.0	1,379	89.8	-	-	KP202303	Silverio et al. 2014	
Colletidae															
Colletinae															
<i>Colletes gigas</i>	15,885	86.2	3,687	85.1	1,344	87.1	836	88.88	1,431	87.7	539	83.5	KM978210	Huang et al. 2014	
Hylaeinae															
<i>Hylaeus dilatatus</i>	15,475	85.8	3,653	84.7	1,309	86.4	802	88.65	1,446	87.1	452	90.9	KP126800	Tan et al. 2015	
Crabronidae															
Philanthinae															
<i>Philanthus triangulum</i>	16,029	83.6	3,707	82.7	1,328	83.6	863	84.2	1,406	86.4	1,039	85.7	JN871914	Kaltenpoth et al. 2012	
Ceraphronoidea															
Megaspilidae															
<i>Conostigmus</i> sp.	16,315	83.0	3,720	81.0	1,370	86.9	785	87.8	1,406	86.6	1447	86.8	KF015227	Mao et al. 2014b	
Ceraphronidae															
<i>Ceraphron</i> sp.	14,947	80.0	3,471	77.9	1,188	85.5	716	86.2	1,349	84.1	692	83.1	KJ570858	Mao et al. 2014a	
Evanioidea															
Evaniidae															
<i>Evania appendigaster</i>	17,817	77.8	3,692	74.4	1,274	79.7	747	75.8	1,445	81.8	2,325	85.6	FJ593187	Wei et al. 2010a	
Gasteruptiidae															
<i>Gasteruption</i> sp.	17,884	82.3	3,737	79.0	1,427	84.9	956	86.8	1,517	85.8	803	89.4	KJ619460	Mao et al. 2014b	
Aulacidae															
<i>Pristaulacus compressus</i>	15,563	84.0	3,731	82.3	1,295	86.0	813	88.7	1,501	87.1	780	91.9	KF500406	Wei et al. 2015a	
Ichneumonoidea															
Ichneumonidae															
Campopleginae															
<i>Diadegma semiclausum</i>	18,728	87.4	3,696	83.7	1,392	88.0	768	88.5	1,506	87.3	2,161	96.4	EU871947	Wei et al. 2009	
Doryctinae															
<i>Spathius agrili</i>	15,425	84.0	3,741	82.1	1,303	88.8	749	8/9.3	1,476	87.1	578	93.6	FJ387020	Wei et al. 2010b	
Microgastrinae															
<i>Cotesia vestalis</i>	15,543	87.2	3,715	85.7	1,274	89.7	764	91.2	1,525	90.0	571	91.6	FJ154897	Wei et al. 2010b	
Ichneumoninae															

<i>Diadromus collaris</i>	14,621	84.0	3,686	82.8	1,371	87.4	624	85.4	1,298	87.7	-	-	JX131613	Li et al. 2015
Ophioninae														
<i>Enicospilus</i> sp.	15,300	84.0	3,718	84.0	1,294	87.2	861	89.4	1,380	87.5	-	-	FJ478177	Dowton et al. 2009
Platygastroidea														
Scelionidae														
Scelioninae														
<i>Ceratobaeus</i> sp.	15,851	77.6	3,690	75.5	1,280	84.1	754	84.4	1,438	85.5	1,099	75.2	KF696669	Mao and Dowton 2014
<i>Idris</i> sp.	18,996	81.2	3,683	79.0	1,289	87.4	774	86.3	1,446	87.6	414	84.3	KF696670	Mao and Dowton 2014
Telenominae														
<i>Trissolcus basalis</i>	15,768	84.1	3,699	82.2	1,270	88.7	752	86.2	1,396	88.7	-	-	JN903532	Mao et al. 2012
Proctotrupoidea														
Pelecinidae														
<i>Pelecinus polyturator</i>	14,896	81.2	3,668	79.2	1,271	86.3	734	86.9	1,468	86.1	192	95.3	KM104167	Mao et al. 2015
Vanhorniidae														
<i>Vanhornia eucnemidarum</i>	16,574	80.1	3,678	78.1	1,327	82.8	395	80.8	2,149	84.3	-	-	DQ302100	Castro et al. 2006
Symphyta														
Tenthredinoidea														
Tenthredinidae														
Allantinae														
<i>Allantus luctifer</i>	15,418	81.1	3,725	79.8	1,340	84.1	807	83.5	1,484	83.8	807	83.5	KJ713152	Wei et al. 2014c
Nematinae														
<i>Monocellicampa pruni</i>	15,169	77.2	3,713	75.1	1,356	82.7	797	81.2	1,293	82.5	-	-	JX566509	Wei et al. 2015b

^aTermination codons were excluded in total codon count. ^bProtein coding genes. Asterisk (*) indicates incomplete mitochondrial genome sequences; and bar (-) indicates unavailable sequences in the genome.

Table 4. Frequency of four most frequently used codons in Hymenoptera.

Group	Species	Codon				Total
		ATT (I)	TTA (L)	TTT (F)	ATA (M)	
Apoidea	<i>Apis florea</i>	12.79	14.07	9.79	8.40	45.05
Apoidea	<i>Bombus ignitus</i>	13.55	13.55	10.10	9.63	46.83
Apoidea	<i>Colletes gigas</i>	12.99	14.24	10.14	10.09	47.46
Apoidea	<i>Hylaeus dilatatus</i>	12.81	13.50	9.99	11.17	47.47
Apoidea	<i>Melipona scutellaris</i>	13.12	13.42	12.06	10.70	49.30
Apoidea	<i>Philanthus triangulum</i>	11.93	13.36	9.85	9.31	44.44
Ceraphronoidea	<i>Ceraphron</i> sp.	10.75	11.74	9.48	8.49	40.45
Ceraphronoidea	<i>Conostigmus</i> sp.	11.34	12.04	8.82	9.84	42.04
Evanoidea	<i>Evania appendigaster</i>	10.08	8.23	8.02	7.64	33.97
Evanoidea	<i>Gasteruption</i> sp.	9.85	11.27	9.02	8.54	38.67
Evanoidea	<i>Pristaulacus compressus</i>	13.19	12.73	11.55	7.40	44.87
Formicidae	<i>Atta laevigata</i>	10.62	9.72	9.24	8.32	37.89
Formicidae	<i>Camponotus atrox</i>	9.91	9.74	8.95	7.70	36.30
Formicidae	<i>Formica fusca</i>	12.15	11.86	9.90	8.36	42.27
Formicidae	<i>Leptomyrmex pallens</i>	6.67	5.97	7.84	6.86	27.34
Formicidae	<i>Myrmica scabrinodis</i>	12.41	11.18	9.66	8.73	41.98
Formicidae	<i>Pristomyrmex punctatus</i> (normal)	10.87	9.76	8.78	8.64	38.05
Formicidae	<i>Solenopsis richteri</i>	9.86	8.09	8.61	6.94	33.50
Vespoidea	<i>Abispa ephippium</i>	10.68	10.52	8.87	8.84	38.90
Vespoidea	<i>Vespa bicolor</i>	11.23	11.36	9.22	8.47	40.28
Vespoidea	<i>Wallacidia oculata</i>	8.82	8.28	7.58	9.92	34.61
Ichneumonoidea	<i>Cotesia vestalis</i>	12.46	14.59	12.52	9.56	49.13
Ichneumonoidea	<i>Diadegma semiclausum</i>	13.18	13.53	11.12	8.66	46.48
Ichneumonoidea	<i>Diadromus collaris</i>	13.65	13.35	9.79	9.88	46.66
Ichneumonoidea	<i>Enicospilus</i> sp.	13.39	14.07	9.76	9.84	47.07
Ichneumonoidea	<i>Spathius agrili</i>	10.41	14.09	10.46	8.28	43.24
Platygastroidea	<i>Ceratobaeus</i> sp.	10.08	9.05	8.08	8.38	35.59
Platygastroidea	<i>Idris</i> sp.	11.27	11.11	10.18	7.96	40.51
Platygastroidea	<i>Trissolcus basalis</i>	12.57	12.44	11.00	9.22	45.23
Proctotrupoidea	<i>Pelecinus polyturator</i>	10.52	12.19	9.68	8.59	40.98
Proctotrupoidea	<i>Vanhornia eucnemidarum</i>	9.96	9.25	7.35	8.95	35.52
Tenthredinoidea	<i>Allantus luctifer</i>	11.14	12.70	10.15	7.65	41.64
Tenthredinoidea	<i>Monocellicampa pruni</i>	9.24	11.02	8.62	7.22	36.09

Within parenthesis indicate corresponding amino acid.

Fig. 1.

(A) *Camponotus atrox*

(B) *Formica fusca*

(C) *Formica selysi*

(D) *Pristomyrmex punctatus*

Fig. 2.

Fig. 3.

184x140mm (300 x 300 DPI)

Fig. 4.

122x102mm (300 x 300 DPI)

Ancestral type for insects

Formicinae

Camponotus atrox
C. chromaiodes^a
Polyrhachis dives^b

Dolichoderinae

Leptomyrmex pallens

Formicinae

Formica fusca
F. selysi

Myrmicinae

Solenopsis richteri
S. geminate, *S. invicta*
Atta laevigata^c
Myrmica scabrinodis

Myrmicinae

Pristomyrmex punctatus

Fig. 5.

Fig. 6.

gen-2015-0080.R1 Supplement Table 1. Gene sizes (bp) of ant mitochondrial genomes.

Gene	Anticodon	Start codon	Stop codon	<i>C. atrox</i>	<i>C. chromaiodes</i>	<i>P. dives</i>	<i>F. fusca</i>	<i>F. selysi</i>	<i>S. richteri</i>	<i>S. geminata</i>
Size				16,540	12,473	13,129	16,673	16,728	15,560	15,552
<i>trnM</i>	CAT			70	-	-	69	69	68	69
<i>trnI</i>	GAT			70	-	-	68	68	70	67
<i>trnQ</i>	TTG			71	-	-	70	70	69	69
<i>ND2</i>		ATT	TAA ¹ /T ²	999 ¹	984 ¹	-	1023 ¹	1023	982 ¹	982 ²
<i>trnW</i>	TCA			73	74*	69	73	73	72	70
<i>trnC</i>	GCA			66	66*	68	69	69	68	71
<i>trnY</i>	GTA			66	66*	66	67	67	66	66
<i>COI</i>		ATG	TAA ¹ /T ²	1530 ¹	1530 ¹	1530 ¹	1530 ¹	1530 ¹	1528 ²	1528 ²
<i>trnL₂</i>	TAA			69	69*	72	72	72	68	67
<i>COII</i>		ATT ^a /ATA ^b	TAA	687 ^a	687 ^a	678 ^a	681 ^a	681 ^a	690 ^b	690 ^a
<i>trnK</i>	TTT			70	70*	72	71	71	75	73
<i>trnD</i>	GTC			69	69*	69	75	75	72	70
<i>ATP8</i>		ATA ^a /ATT ^b	TAA ¹ /T ²	156 ^{a,1}	156 ^{a,1}	162 ^{b,1}	168 ^{b,1}	168 ^{b,1}	154 ^{a,2}	154 ^{a,2}
<i>ATP6</i>		ATG ^a /ATA ^b /ATC ^c	TAA ¹ /TAG ²	669 ^{a,1}	669 ^{a,1}	666 ^{b,1}	669 ^a	669 ^{a,1}	666 ^{b,1}	666 ^{b,1}
<i>COIII</i>		ATG ^a /ATA ^b	TAA	792 ^a	792 ^a	792 ^a	795 ^b	795 ^b	783 ^a	783 ^a
<i>trnG</i>	TCC			67	67*	72	72	72	69	69
<i>ND3</i>		ATA ^a /ATT ^b /ATC ^c	TAA ¹ /TAG ²	354 ^{a,1}	354 ^{a,1}	351 ^{a,1}	351 ^{b,1}	351 ^{b,1}	351 ^{a,1}	351 ^{c,1}
<i>trnA</i>	TGC			65	65*	65	75	75	69	66
<i>trnR</i>	TCG			71	71*	51	67	67	67	66
<i>trnN</i>	GTT			68	68*	66	73	73	67*	68*
<i>trnS₁</i>	TCT			58	58*	59	64	64	57	59
<i>trnE</i>	TTC			68	69*	60	74	74	67	66
<i>trnF</i>	GAA			66	67*	68	70	67	68	68
<i>ND5</i>		ATA ^a /ATT ^b /ATG ^c	TAA	1653 ^a	1668 ^a	1665 ^a	1671 ^b	1671 ^b	1668 ^b	1668 ^b
<i>trnH</i>	GTG			71	67*	69	65	65	67	68
<i>ND4</i>		ATT ^a /ATG ^b /ATA ^c	TAA ¹ /TAG ²	1341 ^{a,1}	1329 ^{b,1}	1326 ^{b,1}	1347 ^{a,1}	1347 ^{a,1}	1338 ^{a,1}	1338 ^{a,1}
<i>ND4L</i>		ATA ^a /ATT ^b	TAA ¹ /TAG ²	297 ^{b,1}	285 ^{a,1}	339 ^{a,2}	294 ^{a,1}	288 ^{b,1}	288 ^{b,2}	288 ^{b,1}
<i>trnT</i>	TGT			65	65*	78	62	62	65	64
<i>trnP</i>	TGG			95	88*	74	67	67	69	69
<i>ND6</i>		ATG ^a /ATT ^b /ATA ^c	TAA	549 ^a	549 ^a	540 ^c	552 ^a	552 ^a	531 ^a	531 ^a
<i>CytB</i>		ATG ^a /ATA ^b	TAA ¹ /TA ²	1116 ^{a,1}	1115 ^{a,1}	798 ^{b,1}	1125 ^{a,1}	1125 ^{a,1}	1122 ^{a,1}	1122 ^{a,1}
<i>trnS₂</i>	TGA			69	69*	72	74	74	68	68
<i>ND1</i>		ATA ^a /TTG ^b /ATT ^c	TAA ¹ /T ² /TAG ³	960 ^{a,1}	960 ^{a,1}	957 ^{a,1}	960 ^{a,1}	960 ^{c,1}	946 ^{a,2}	949 ^{c,2}
<i>trnL₁</i>	TAG			90	-	55	83	85	79	79
<i>lrRNA</i>				1428	-	203	1502	1520	1333	1338
<i>trnV</i>	TAC			64	-	-	82	82	66	67
<i>srRNA</i>				777	-	-	802	805	784	783
A+T-rich region				1402	-	-	529	585	379*	376

Gene	Anticodon	Start codon	Stop codon	<i>S. invicta</i> (PMS)	<i>S. invicta</i> (VMS)	<i>A. laevigata</i>	<i>P. punctatus</i> (normal)	<i>P. punctatus</i> (cheaters)	<i>M. scabrinodis</i>	<i>L. pallens</i>
Size				15,549	15,548	18,729	16,269	16,180	15,310	15,591
<i>trnM</i>	CAT			67	69	71	69	69	68	67
<i>trnI</i>	GAT			70	70	72	72	72	66	66
<i>trnQ</i>	TTG			69	69	70	70	70	69	71
ND2		ATT	TAA ¹ /T ²	982 ²	982 ²	981 ¹	993 ¹	993 ¹	976 ²	999 ²
<i>trnW</i>	TCA			72	73	70	70	70	69	70
<i>trnC</i>	GCA			66	68	70	72	72	66	63
<i>trnY</i>	GTA			67	65	66	63	63	71	66
<i>COI</i>		ATG	TAA ¹ /T ²	1528 ²	1528 ²	1533 ¹	1533 ¹	1533 ¹	1530 ¹	1533 ¹
<i>trnL₂</i>	TAA			68	68	71	69	69	69	67
<i>COII</i>		ATT ^a /ATA ^b	TAA	690 ^a	690 ^a	684 ^a	684 ^a	684 ^a	693 ^a	717 ^a
<i>trnK</i>	TTT			75	75	70	72	72	70	70
<i>trnD</i>	GTC			72	73	69	68	68	66	67
<i>ATP8</i>		ATA ^a /ATT ^b	TAA ¹ /T ²	154 ^{a,2}	159 ^{a,1}	189 ^{a,1}	174 ^{a,1}	174 ^{a,1}	159 ^{b,1}	153 ^{b,1}
<i>ATP6</i>		ATG ^a /ATA ^b /ATC ^c	TAA ¹ /TAG ²	666 ^{b,1}	666 ^{b,1}	666 ^{b,2}	672 ^{c,1}	672 ^{c,1}	666 ^{a,1}	663 ^{b,1}
<i>COIII</i>		ATG	TAA	783 ^a	783 ^a	792 ^a	783 ^a	783 ^a	783 ^a	801 ^a
<i>trnG</i>	TCC			72	74	65	70	70	66	69
ND3		ATA ^a /ATT ^b /ATC ^c	TAA ¹ /TAG ²	351 ^{a,1}	351 ^{a,2}	354 ^{b,1}	351 ^{b,1}	351 ^{b,1}	351 ^{b,1}	354 ^{a,1}
<i>trnA</i>	TGC			69	67	66	70	70	72	67
<i>trnR</i>	TCG			67	67	74	59	59	65	68
<i>trnN</i>	GTT			68*	67	70	65	65	66	65*
<i>trnS_I</i>	TCT			57	57	61	60	60	59	59
<i>trnE</i>	TTC			67	67	68	70	70	68	72
<i>trnF</i>	GAA			68	68	69	68	68	67	65
ND5		ATA ^a /ATT ^b /ATG ^c	TAA	1668 ^b	1668 ^b	1665 ^b	1665 ^c	1665 ^c	1659 ^a	1656 ^a
<i>trnH</i>	GTG			67	67	73	66	66	69	71
ND4		ATT ^a /ATG ^b /ATA ^c	TAA ¹ /TAG ²	1338 ^{a,1}	1338 ^{a,1}	1347 ^{c,2}	1329 ^{b,1}	1329 ^{b,1}	1323 ^{b,2}	1347 ^{b,1}
ND4L		ATA ^a /ATT ^b	TAA ¹ /TAG ²	288 ^{b,2}	288 ^{b,2}	276 ^{b,2}	276 ^{b,1}	276 ^{b,1}	289 ^{b,1}	288 ^{a,1}
<i>trnT</i>	TGT			65	64	70	67	67	66	64
<i>trnP</i>	TGG			69	69	73	64	64	66	72
ND6		ATG ^a /ATT ^b /ATA ^c	TAA	531 ^a	531 ^a	561 ^a	552 ^b	552 ^b	537 ^a	540 ^a
<i>CytB</i>		ATG ^a /ATA ^b	TAA ¹ /TA ²	1122 ^{a,1}	1122 ^{a,1}	1119 ^{a,1}	1104 ^{b,1}	1104 ^{b,1}	1112 ^{a,2}	1122 ^{a,1}
<i>trnS₂</i>	TGA			68	68	70	71	71	69	68
ND1		ATA ^a /TTG ^b /ATT ^c	TAA ¹ /T ² /TAG ³	946 ^{a,2}	946 ^{a,2}	948 ^{a,1}	960 ^{a,1}	960 ^{a,1}	951 ^{c,3}	936 ^{b,3}
<i>trnL_I</i>	TAG			79	78	71	61	61	68	68
<i>lrRNA</i>				1333	1332	1426	1330	1327	1296	1274
<i>trnV</i>	TAC			66	66	69	74	74	66	68
<i>srRNA</i>				781	784	795	800	800	784	774
A+T-rich region				377	506	-	966	881	472	195

Note: -, no sequence available. Identical superscripts indicate identical start codon and stop codon for corresponding genes. *Reannotated sequences using GenBank data. The abbreviations for the genes are as follows: *COI*, *COII* and *COIII* refer to the cytochrome oxidase subunits, *CytB* refers to cytochrome b, and *ND1-6* refers to NADH dehydrogenase components. *srRNA* and *lrRNA* denote small and large subunit ribosomal RNA (rRNA) genes. tRNAs are denoted as one-letter symbols in accordance with the IUPAC-IUB single-

letter amino acid codes except those encoding leucine and serine, which are labeled L_1 for CTN codon family, L_2 for TTR codon family, S_1 for AGN codon family, and S_2 for TCN codon family.

Draft

gen-2015-0080.R1 Supplement Table 2. Overlapping and intergenic-spacer sequences of ant mitochondrial genomes.

Region	A	B	C	D	E	F	G	H	I	J	K	L	M	N
<i>trnI</i> - <i>trnM</i>	(I)11	-												
<i>trnM</i> - <i>trnQ</i>	(I)57	-												
<i>trnV</i> - <i>trnM</i>						(I)61	(I)58	(I)61	(I)60	(I)101	(I)78	(I)78	(I)84	
<i>trnM</i> - <i>trnI</i>		-		(I)100	(I)103	(I)17	(I)19	(I)18	(I)19	(I)166	(I)7	(I)7	(I)18	
<i>trnI</i> - <i>trnQ</i>		-		(I)15	(I)15	(I)13	(I)13	(I)13	(I)13	(I)93	(I)1	(I)1	(I)9	(I)6
<i>trnQ</i> - <i>ND2</i>	(I)66	-		(I)65	(I)65	(I)86	(I)78	(I)86	(I)86	(I)189	(I)54	(I)54	(I)50	(I)145
<i>ND2</i> - <i>trnW</i>	(I)3	(I)3	-	(I)5	(I)5	(I)3	(I)3	(I)3	(I)3	(I)8	(I)62	(I)62		(I)6
<i>trnW</i> - <i>trnC</i>	(O)1	(O)1	(I)20	(O)1	(O)1	(O)1	(O)1	(O)1	(O)1	(I)11	(I)4	(I)4	(I)11	(I)52
<i>trnC</i> - <i>trnY</i>	(I)2	(I)2	(I)36	(I)18	(I)18	(I)6	(I)5	(I)6	(I)5	(I)118	(I)14	(I)14	(O)1	(I)8
<i>trnY</i> - <i>COI</i>	(I)10	(I)20	(I)18	(I)24	(I)24	(I)13	(I)11	(I)13	(I)12	(I)202	(I)11	(I)11	(I)11	(I)29
<i>COI</i> - <i>trnL₂</i>	(I)20	(I)20	(I)130	(I)26	(I)28	(I)1	(I)1	(I)1	(I)1	(I)160	(O)5	(O)5	(I)9	(O)4
<i>trnL₂</i> - <i>COII</i>														(O)1
<i>COII</i> - <i>trnK</i>	(I)12	(I)19	(I)19	(I)78	(I)78	(I)12	(I)4	(I)13	(I)6	(I)196			(I)3	(I)81
<i>COII</i> - <i>trnD</i>											(I)27	(I)27		
<i>trnK</i> - <i>trnD</i>				(O)1	(O)1	(I)13	(I)11	(I)13	(I)12	(I)236			(O)1	(O)2
<i>trnD</i> - <i>trnK</i>											(I)31	(I)31		
<i>trnD</i> - <i>ATP8</i>										(I)167				(O)1
<i>trnK</i> - <i>ATP8</i>											(I)66	(I)66		
<i>ATP8</i> - <i>ATP6</i>	(O)7	(O)7	(O)4	(O)7	(O)7	(I)1	(I)1	(I)1	(O)4	(O)4	(O)10	(O)10	(O)7	(O)4
<i>ATP6</i> - <i>COIII</i>	(I)27	(I)34	(I)15	(I)38	(I)38	(I)7	(I)7	(I)7	(I)7	(I)91	(I)8	(I)8	(I)3	(I)5
<i>COIII</i> - <i>trnG</i>	(I)17	(I)29	(I)24	(I)40	(I)40	(I)13	(I)6	(I)13	(I)6	(I)215	(I)27	(I)27	(I)4	(I)20
<i>trnG</i> - <i>ND3</i>	(O)3	(O)3												(O)4
<i>ND3</i> - <i>trnA</i>	(I)10	(I)10	(I)10	(I)22	(I)21	(I)8	(I)8	(I)8	(I)8	(I)57	(I)33	(I)33	(I)4	(I)7
<i>trnA</i> - <i>trnR</i>			(I)3	(I)17	(I)17	(I)28	(I)32	(I)25	(I)25	(I)85	(I)15	(I)15	(I)15	(I)4
<i>trnR</i> - <i>trnN</i>	(O)11	(O)11	(I)2	(I)31	(I)33	(I)3	(I)3	(I)3	(I)3	(I)207	(I)5	(I)5		(I)6
<i>trnN</i> - <i>trnS₁</i>	(I)7	(I)7	(I)75	(I)31	(I)31	(O)3	(O)3	(O)3	(O)3	(O)3	(O)4	(O)4	(I)11	(I)1
<i>trnS₁</i> - <i>trnE</i>	(I)2	(I)2		(I)75	(I)74	(I)1	(I)1	(I)1	(I)1	(O)1	(I)23	(I)23		(I)8
<i>trnE</i> - <i>trnF</i>	(I)8	(I)8	(I)24	(O)12	(O)8	(O)8	(O)8	(O)8	(O)8	(O)8	(O)8	(O)8	(O)8	(O)2
<i>trnF</i> - <i>ND5</i>	(I)2	(I)2	(I)2	(I)81	(I)83	(O)1	(O)1	(O)1	(O)1	(I)13	(I)3	(I)3	(I)2	(I)4
<i>ND5</i> - <i>trnH</i>	(I)11	(O)3	(I)9											(I)1
<i>trnH</i> - <i>ND4</i>	(I)4	(I)4	(I)10	(I)10	(I)10	(I)12	(I)12	(I)12	(I)12	(I)8	(I)11	(I)11	(I)1	(I)28
<i>ND4</i> - <i>ND4L</i>	(O)1	(I)13	(I)11	(I)29	(I)29	(O)7	(O)7	(O)7	(O)7	(I)247	(I)9	(I)9	(I)1	(I)13
<i>ND4L</i> - <i>trnT</i>	(I)1	(I)13	(O)49	(I)59	(I)65	(I)1	(I)1	(I)1	(I)1	(I)11	(I)18	(I)18	(I)1	(I)2
<i>trnT</i> - <i>trnP</i>	(I)11	(I)11	(I)23	(I)15	(I)18	(O)1	(O)1	(O)1	(O)1	(I)1	(I)5	(I)5	(I)2	(I)15
<i>trnP</i> - <i>ND6</i>	(I)12	(I)12	(I)11	(I)23	(I)18	(I)41	(I)38	(I)37	(I)41	(I)84	(I)24	(I)24	(I)1	(I)38
<i>ND6</i> - <i>CytB</i>	(I)5	(I)5	(I)361	(I)43	(I)38	(I)24	(I)34	(I)24	(I)24	(I)70	(I)10	(I)10	(I)3	(I)4
<i>CytB</i> - <i>trnS₂</i>	(I)8	(I)8	(I)17	(I)59	(I)59	(I)12	(I)12	(I)12	(I)12	(I)257	(I)75	(I)75		(I)12
<i>trnS₂</i> - <i>ND1</i>	(I)9	(I)29	(I)81	(I)16	(I)16	(I)6	(I)6	(I)6	(I)7	(I)322	(I)12	(I)12	(I)6	(I)6
<i>ND1</i> - <i>trnL₁</i>	(O)3	-	(I)36			(I)3		(I)3	(I)3	(I)176	(I)10	(I)10		(I)15
<i>trnL₁</i> - <i>lrRNA</i>		-	(I)66	(I)125	(I)125	(I)12	(I)15	(I)12	(I)12	(I)221	(I)26	(I)26		(I)12
<i>lrRNA</i> - <i>trnV</i>		-	-	(I)66	(I)68									(I)33
<i>trnV</i> - <i>srRNA</i>		-	-	(I)22	(I)19									(O)26

<i>lrRNA</i> - <i>srRNA</i>						(I)75	(I)90	(I)75	(I)79	(I)95	(O)35	(O)35	(I)17	
<i>srRNA</i> - A+T-rich	-					(I)57	(I)57	(I)57	(I)57	-				(I)232
<i>srRNA</i> - <i>trnV</i>										(I)74				
A+T-rich region - <i>trnM</i>														(I)68
Total nucleotides	(O)26	(O)25	(O)53	(O)21	(O)17	(O)21	(O)21	(O)21	(O)25	(O)16	(O)62	(O)63	(O)17	(O)44
	(I)315	(I)251	(I)1003	(I)1135	(I)1138	(I)529	(I)526	(I)524	(I)516	(I)3881	(I)669	(I)669	(I)266	(I)793

Note: (O), Overlapping sequences; and (I), Intergenic space sequences. Empty column means that neighboring genes are abutting to each other.
A, *C. atrox*; B, *C. chromaiodes*; C, *P. dives*; D, *F. fusca*; E, *F. selysi*; F, *S. richteri*; G, *S. geminata*; H, *S. invicta* (PMS); I, *S. invicta* (VMS); J, *A. laevigata*; K, *P. punctatus* (normal); L, *P. punctatus* (cheaters); M, *M. scabrinodis*; and N, *L. pallens*.

Draft

gen-2015-0080.R1 Supplement Table 3. Codon position-based nucleotide composition of concatenated 13 PCGs in Hymenoptera.

Group	Species	Overall				1st codon position				2nd codon position				3rd codon position			
		A	T	C	G	A	T	C	G	A	T	C	G	A	T	C	G
Apoidea	<i>Apis florea</i>	37.0	46.8	8.1	8.1	39.6	40.3	8.0	12.1	23.6	52.1	14.0	10.2	47.8	48.1	2.2	1.9
Apoidea	<i>Bombus ignitus</i>	37.6	47.5	7.5	7.4	42.0	39.6	7.6	10.8	24.3	52.7	13.1	9.8	46.4	50.1	1.8	1.7
Apoidea	<i>Colletes gigas</i>	36.8	48.3	7.0	7.9	40.7	40.6	7.3	11.4	24.4	52.0	12.8	10.7	45.2	52.3	0.9	1.6
Apoidea	<i>Hylaeus dilatatus</i>	37.7	47.0	7.7	7.6	42.3	39.2	7.5	11.0	24.1	51.6	13.8	10.5	46.7	50.3	1.8	1.2
Apoidea	<i>Melipona scutellaris</i>	38.2	48.1	7.0	6.7	42.3	41.2	7.2	9.3	24.6	53.9	11.9	9.7	47.6	49.2	1.9	1.2
Apoidea	<i>Philanthus triangulum</i>	36.9	45.8	8.6	8.7	40.3	38.7	8.4	12.6	22.9	51.5	14.3	11.3	47.5	47.2	3.1	2.2
Ceraphronoidea	<i>Ceraphron</i> sp.	33.2	44.9	10.9	11.0	38.6	37.0	10.1	14.4	20.7	51.4	15.0	12.9	40.4	46.3	7.5	5.8
Ceraphronoidea	<i>Conostigmus</i> sp.	35.9	45.1	9.9	9.1	40.7	37.4	9.3	12.7	22.6	51.6	14.8	11.1	44.4	46.4	5.6	3.5
Evanioidea	<i>Evania appendigaster</i>	31.7	42.7	13.2	12.4	36.5	34.0	12.9	16.6	20.2	49.7	16.8	13.4	38.4	44.5	9.9	7.2
Evanioidea	<i>Gasteruption</i> sp.	33.7	45.2	9.5	11.6	37.8	37.8	8.8	15.5	22.6	50.8	14.4	12.2	40.8	47.0	5.2	7.0
Evanioidea	<i>Pristaulacus compressus</i>	35.0	47.4	9.3	8.3	39.3	40.3	9.2	11.2	21.9	52.6	14.3	11.1	43.7	49.2	4.4	2.7
Formicidae	<i>Atta laevigata</i>	32.9	44.7	12.1	10.2	38.1	36.3	11.5	14.1	21.4	50.9	16.0	11.8	39.2	47.0	9.0	4.8
Formicidae	<i>Camponotus atrox</i>	32.2	44.1	12.8	10.9	37.6	37.8	11.4	13.3	22.3	51.5	14.9	11.3	36.8	43.0	12.0	8.1
Formicidae	<i>Formica fusca</i>	34.3	46.8	10.0	8.9	38.1	39.1	10.0	12.8	22.6	51.3	15.0	11.1	42.1	50.0	4.9	3.0
Formicidae	<i>Leptomyrmex pallens</i>	28.6	38.7	17.5	15.2	34.5	30.8	16.1	18.6	19.1	48.9	18.2	13.8	32.2	36.5	18.3	13.0
Formicidae	<i>Myrmica scabrinodis</i>	33.9	46.2	10.2	9.6	39.6	37.5	10.2	12.7	22.5	51.1	14.8	11.7	39.8	50.0	5.7	4.5
Formicidae	<i>Pristomyrmex punctatus</i> (normal)	33.9	44.0	11.9	10.2	39.5	36.2	10.9	13.4	21.7	50.8	15.8	11.7	40.4	45.0	9.1	5.5
Formicidae	<i>Solenopsis richteri</i>	31.2	42.8	14.4	11.6	37.0	35.3	13.2	14.5	21.4	50.4	16.0	12.1	35.1	42.8	14.0	8.2
Vespoidea	<i>Abispa ephippium</i>	35.1	43.5	11.2	10.1	38.4	35.3	11.5	14.8	22.2	50.6	15.3	11.9	44.7	44.6	6.9	3.7
Vespoidea	<i>Vespa bicolor</i>	34.9	44.3	11.1	9.6	39.1	36.3	11.2	13.3	22.1	51.1	15.2	11.5	43.4	45.6	7.0	4.0
Vespoidea	<i>Wallacidia oculata</i>	33.8	41.1	12.4	12.6	38.9	33.3	10.7	17.0	20.0	49.5	17.0	13.4	42.5	40.6	9.5	7.4
Ichneumonoidea	<i>Cotesia vestalis</i>	35.5	50.2	6.5	7.9	39.7	42.1	6.7	11.5	23.7	53.8	12.0	10.4	43.0	54.6	0.8	1.6
Ichneumonoidea	<i>Diadegma semiclausum</i>	36.7	47.0	8.2	8.1	40.7	39.7	8.1	11.5	23.5	51.5	13.7	11.2	45.8	49.8	2.7	1.7
Ichneumonoidea	<i>Diadromus collaris</i>	36.6	46.2	9.0	8.2	41.4	38.3	8.8	11.5	22.6	51.6	14.5	11.3	45.9	48.7	3.6	1.8
Ichneumonoidea	<i>Enicospilus</i> sp.	37.3	46.7	7.7	8.3	40.9	39.0	8.0	12.0	23.7	52.0	13.3	11.1	47.1	49.2	1.9	1.7
Ichneumonoidea	<i>Spathius agrili</i>	34.9	47.1	8.0	10.0	37.3	41.1	7.9	13.7	23.8	51.3	13.0	12.0	43.6	48.9	3.2	4.2
Platygastridae	<i>Ceratobaeus</i> sp.	33.6	41.9	13.1	11.4	40.3	34.2	11.5	13.9	22.2	51.6	14.7	11.4	38.2	39.8	13.2	8.8
Platygastridae	<i>Idris</i> sp.	33.9	45.1	11.0	10.0	39.5	37.1	10.3	13.1	22.4	52.5	13.6	11.5	39.8	45.8	9.0	5.5
Platygastridae	<i>Trissolcus basalidis</i>	36.5	45.7	9.1	8.7	40.8	38.0	8.7	12.4	23.5	51.9	14.0	10.6	45.0	47.3	4.8	2.9
Proctotrupoidea	<i>Pelecinus polyturator</i>	34.3	44.9	11.2	9.6	38.8	38.4	10.2	12.5	22.3	51.4	15.1	11.2	41.8	44.9	8.2	5.1
Proctotrupoidea	<i>Vanhornia eucnemidarum</i>	35.4	42.7	11.8	10.1	40.2	36.0	10.9	12.8	22.3	49.4	16.2	12.0	43.7	42.7	8.2	5.4
Tenthredinoidea	<i>Allantus luctifer</i>	34.8	45.0	10.2	10.0	37.5	37.8	9.9	14.8	21.4	49.4	16.6	12.6	45.6	47.8	4.1	2.5
Tenthredinoidea	<i>Monocellicampa pruni</i>	32.9	42.2	13.0	11.9	36.6	36.0	11.6	15.8	21.1	48.0	17.6	13.3	41.0	42.5	9.8	6.7

Stop codon was excluded in the count.

gen-2015-0080.R1 Supplement Table 4. Skewness of concatenated 13PCGs in Hymenoptera

Group	Species	Whole PCGs		1 st codon position		2 nd codon position		3 rd codon position	
		A/T	G/C	A/T	G/C	A/T	G/C	A/T	G/C
Apoidea	<i>Apis florea</i>	-0.117	0.000	-0.009	0.201	-0.377	-0.157	-0.003	-0.053
Apoidea	<i>Bombus ignitus</i>	-0.117	-0.006	0.029	0.173	-0.369	-0.143	-0.038	-0.039
Apoidea	<i>Colletes gigas</i>	-0.136	0.063	0.001	0.222	-0.361	-0.090	-0.073	0.311
Apoidea	<i>Hylaeus dilatatus</i>	-0.110	-0.008	0.038	0.190	-0.364	-0.133	-0.037	-0.207
Apoidea	<i>Melipona scutellaris</i>	-0.115	-0.019	0.012	0.131	-0.374	-0.100	-0.016	-0.246
Apoidea	<i>Philanthus triangulum</i>	-0.108	0.006	0.020	0.200	-0.384	-0.117	0.003	-0.170
Ceraphronoidea	<i>Ceraphron</i> sp.	-0.149	0.006	0.022	0.175	-0.426	-0.076	-0.067	-0.130
Ceraphronoidea	<i>Conostigmus</i> sp.	-0.114	-0.042	0.043	0.154	-0.390	-0.143	-0.022	-0.226
Evanioidea	<i>Evania appendigaster</i>	-0.148	-0.033	0.035	0.124	-0.422	-0.114	-0.073	-0.161
Evanioidea	<i>Gasteruption</i> sp.	-0.145	0.101	0.001	0.276	-0.385	-0.082	-0.071	0.152
Evanioidea	<i>Pristaulacus compressus</i>	-0.151	-0.055	-0.012	0.097	-0.412	-0.126	-0.059	-0.240
Formicidae	<i>Atta laevigata</i>	-0.153	-0.086	0.023	0.101	-0.409	-0.150	-0.090	-0.303
Formicidae	<i>Camponotus atrox</i>	-0.156	-0.079	-0.003	0.075	-0.396	-0.135	-0.077	-0.194
Formicidae	<i>Formica fusca</i>	-0.155	-0.055	-0.013	0.122	-0.389	-0.152	-0.086	-0.247
Formicidae	<i>Leptomyrmex pallens</i>	-0.150	-0.073	0.057	0.073	-0.439	-0.139	-0.062	-0.168
Formicidae	<i>Myrmica scabrinodis</i>	-0.153	-0.029	0.027	0.111	-0.389	-0.118	-0.114	-0.112
Formicidae	<i>Pristomyrmex punctatus</i> (normal)	-0.130	-0.079	0.044	0.102	-0.401	-0.150	-0.054	-0.245
Formicidae	<i>Solenopsis richteri</i>	-0.158	-0.108	0.023	0.044	-0.404	-0.136	-0.099	-0.263
Vespoidea	<i>Abispa ephippium</i>	-0.107	-0.051	0.043	0.123	-0.391	-0.124	0.001	-0.298
Vespoidea	<i>Vespa bicolor</i>	-0.119	-0.074	0.037	0.085	-0.396	-0.137	-0.024	-0.278
Vespoidea	<i>Wallacidia oculata</i>	-0.098	0.007	0.078	0.226	-0.424	-0.120	0.022	-0.126
Ichneumonoidea	<i>Cotesia vestalis</i>	-0.171	0.095	-0.029	0.266	-0.388	-0.072	-0.119	0.364
Ichneumonoidea	<i>Diadegma semiclausum</i>	-0.123	-0.004	0.012	0.171	-0.373	-0.101	-0.041	-0.239
Ichneumonoidea	<i>Diadromus collaris</i>	-0.116	-0.043	0.038	0.135	-0.392	-0.123	-0.030	-0.327
Ichneumonoidea	<i>Enicospilus</i> sp.	-0.113	0.033	0.024	0.198	-0.374	-0.088	-0.021	-0.067
Ichneumonoidea	<i>Spathius agrili</i>	-0.149	0.109	-0.049	0.270	-0.366	-0.040	-0.057	0.138
Platygastroidea	<i>Ceratobaeus</i> sp.	-0.110	-0.072	0.082	0.095	-0.399	-0.126	-0.020	-0.200
Platygastroidea	<i>Idris</i> sp.	-0.143	-0.045	0.031	0.120	-0.402	-0.084	-0.070	-0.246
Platygastroidea	<i>Trissolcus basalıs</i>	-0.112	-0.028	0.036	0.178	-0.375	-0.136	-0.024	-0.246
Proctotrupoidea	<i>Pelecınus polyturator</i>	-0.134	-0.077	0.005	0.101	-0.395	-0.148	-0.036	-0.233
Proctotrupoidea	<i>Vanhornia eucnemidarum</i>	-0.093	-0.078	0.055	0.080	-0.378	0.149	0.012	0.206
Tenthredinoidea	<i>Allantus luctifer</i>	-0.127	-0.011	-0.003	0.200	-0.396	-0.137	-0.023	-0.239
Tenthredinoidea	<i>Monocellicampa pruni</i>	-0.124	-0.043	0.007	0.156	-0.389	-0.141	-0.018	-0.190

gen-2015-0080.R1 Supplement Table 5. Comparison of genetic divergence of concatenated 13PCGs in Hymenoptera.

Group	Species	GenBank no.	Ka	Ks	Ka/Ks
Apoidea	<i>Apis florea</i>	JX982136	0.47953	7.00832	0.06842
Apoidea	<i>Bombus ignitus</i>	DQ870926	0.51785	6.78011	0.07638
Apoidea	<i>Colletes gigas</i>	KM978210	0.44652	7.86486	0.05677
Apoidea	<i>Hylaeus dilatatus</i>	KP126800	0.43898	7.71124	0.05693
Apoidea	<i>Melipona scutellaris</i>	KP202303	0.53402	6.96562	0.07666
Apoidea	<i>Philanthus triangulum</i>	JN871914	0.41150	7.14972	0.05756
Ceraphronoidea	<i>Ceraphron</i> sp.	KJ570858	0.47994	5.49946	0.08727
Ceraphronoidea	<i>Conostigmus</i> sp.	KF015227	0.47021	8.10356	0.05802
Evanioidea	<i>Evania appendigaster</i>	FJ593187	0.48820	5.81429	0.08397
Evanioidea	<i>Gasteruption</i> sp.	KJ619460	0.39990	7.56064	0.05289
Evanioidea	<i>Pristaulacus compressus</i>	KF500406	0.46343	8.95308	0.05176
Formicidae	<i>Atta laevigata</i>	KC346251	0.44514	7.11500	0.06256
Formicidae	<i>Camponotus atrox</i>	KT159775	0.42621	6.13522	0.06947
Formicidae	<i>Formica fusca</i>	LN607805	0.43369	8.91432	0.04865
Formicidae	<i>Leptomyrmex pallens</i>	KC160533	0.44940	4.57691	0.09819
Formicidae	<i>Myrmica scabrinodis</i>	LN607806	0.42518	8.87390	0.04791
Formicidae	<i>Pristomyrmex punctatus</i> (normal)	AB556947	0.45342	6.91288	0.06559
Formicidae	<i>Solenopsis richteri</i>	HQ215539	0.43424	5.96794	0.07276
Vespoidea	<i>Vespa bicolor</i>	KJ735511	0.45340	5.91554	0.07665
Vespoidea	<i>Abispa ephippium</i>	EU302588	0.45415	5.87116	0.07735
Vespoidea	<i>Wallacidia oculata</i>	FJ611801	0.56365	5.00592	0.11260
Ichneumonoidea	<i>Cotesia vestalis</i>	FJ154897	0.40832	8.52406	0.04790
Ichneumonoidea	<i>Diadegma semiclausum</i>	EU871947	0.38678	8.13545	0.04754
Ichneumonoidea	<i>Diadromus collaris</i>	JX131613	0.36649	8.05313	0.04551
Ichneumonoidea	<i>Enicospilus</i> sp.	FJ478177	0.40893	8.00624	0.05108
Ichneumonoidea	<i>Spathius agrili</i>	FJ387020	0.41295	6.95059	0.05941
Platygastroidea	<i>Ceratobaeus</i> sp.	KF696669	0.48225	5.38345	0.08958
Platygastroidea	<i>Idris</i> sp.	KF696670	0.47035	6.68949	0.07031
Platygastroidea	<i>Trissolcus basalis</i>	JN903532	0.46374	9.14568	0.05071
Proctotrupoidea	<i>Pelecinus polyturator</i>	KM104167	0.40932	7.84184	0.05220
Proctotrupoidea	<i>Vanhornia eucnemidarum</i>	DQ302100	0.40186	5.77227	0.06962
Tenthredinoidea	<i>Allantus luctifer</i>	KJ713152	0.25463	8.30579	0.03066
Tenthredinoidea	<i>Monocellicampa pruni</i>	JX566509	0.27200	7.40730	0.03672

gen-2015-0080.R1 Supplement Legend

gen-2015-0080.R1 Supplement Figure 1. Predicted secondary clover-leaf structures for the 22 tRNA genes of *Camponotus atrox*. tRNA abbreviations follow the IU-PAC-IUB one-letter code. Arms of tRNAs (clockwise from top) indicate the amino acid acceptor (AA) arm, TΨC (T) arm, anticodon (AC) arm, and dihydrouridine (DHU) arm. Gene names that are not underlined indicate a clockwise direction of transcription, whereas those that are underlined are transcribed in a counter-clockwise direction.

gen-2015-0080.R1 Supplement Figure 2. Intergenic spacer sequences that were successfully aligned in ants. Genes are underlined and intergenic spacer sequences are not underlined. Asterisks indicate identical sequences in the alignment. The percentage of sequence homology in the alignment is presented to the right of the alignment. Bar (-) indicates a deletion with respect to other sequences in the alignment.

<i>trnQ</i> - <i>ND2</i>		<i>trnQ</i>	<i>ND2</i>	
Formicinae, Camponotini	<i>C. atrox</i>	<u>TTAAAAATATATTTATTTTTAATTTTTTAAATTAAA--AATTTTTTAATATTTTTTTAATTATTATAATTATTATTAATTA</u>	<u>ATTA</u>	66.1 %
Formicinae, Formicini	<i>F. fusca</i>	<u>TTAAAAATATATTTAAAAATAAATAATTTTTAATTAAATAAATATATTTTTATTATAATTAATTTTTTAAATTATTA--ATTTTATATA</u>		
		***** *		
<i>COI</i> - <i>trnL₂</i>		<i>COI</i>	<i>trnL₂</i>	
Formicinae, Camponotini	<i>C. atrox</i>	<u>CTCAATTTAAACTATTAATTAATTTAAATTTTAAATATGG</u>		85.0 %
Formicinae, Camponotini	<i>C. chromaiodes</i>	<u>CTCAATTTAAATTATTAATCAATCTAAATTTTAAATATGG</u>		
		***** * * * *		
<i>ATP6</i> - <i>COIII</i>		<i>ATP6</i>	<i>COIII</i>	
Formicinae, Camponotini	<i>C. atrox</i>	<u>AGAAATTTAATAATAATTA----TAATTAT---TATATTCACCTTATGATAAAAA</u>		80.0 %
Formicinae, Camponotini	<i>C. chromaiodes</i>	<u>GGAAATTTAATAATAATTAATAATAATTATAATTATATTCATTTATGATAAAAA</u>		
		***** *		
<i>COIII</i> - <i>trnG</i>		<i>COIII</i>	<i>trnG</i>	
Formicinae, Camponotini	<i>C. atrox</i>	<u>AAGATATTAATCAAAACAATTAATCT-----TATTTATATAG</u>		53.5 %
Formicinae, Camponotini	<i>C. chromaiodes</i>	<u>AAGATATTAATTTAAACAATTAATCCCCCAACCAATAATATTTATATAG</u>		
		***** *		
<i>ND3</i> - <i>trnA</i>		<i>ND3</i>	<i>trnA</i>	
Formicinae, Camponotini	<i>C. atrox</i>	<u>AAAAATTTAACAAATATATAAGGATTTTAG</u>		90.0 %
Formicinae, Camponotini	<i>C. chromaiodes</i>	<u>AAAAATTTAATAAATATATAAGGATTTTAG</u>		
		***** *		
<i>trnT</i> - <i>trnP</i>		<i>trnT</i>	<i>trnP</i>	
Formicinae, Camponotini	<i>C. atrox</i>	<u>ATCATATGATAAAATTATACATCAAAAAATA</u>		90.9 %
Formicinae, Camponotini	<i>C. chromaiodes</i>	<u>ATCATATGATAAAATTATACGTCAAAAATA</u>		
		***** *		
<i>trnP</i> - <i>ND6</i>		<i>trnP</i>	<i>ND6</i>	
Formicinae, Camponotini	<i>C. atrox</i>	<u>CTATTTTTTGCTTTAATAACTTATGATAAAAT</u>		83.3 %
Formicinae, Camponotini	<i>C. chromaiodes</i>	<u>CTATTTTTGTTTCAATAACTTATGATAAAAT</u>		
		***** *		

<https://mc06.manuscriptcentral.com/genome-pubs>

ND3 - trnA

Formicinae, Formicini	<i>F. fusca</i>	<u>AAAATATTAAAATATAAAGTAAATTTTATTTTAGGATTTTAG</u>	90.9 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>AAAATATTAAA-TATAAAATAAATTTTATTTTAGGATTTTAG</u>	

trnA - trnR

Formicinae, Formicini	<i>F. fusca</i>	<u>ATAAATCTTATAAATTATTTTATACATTCTTTTAAAG</u>	100 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>ATAAATCTTATAAATTATTTTATACATTCTTTTAAAG</u>	

trnR - trnN

Formicinae, Formicini	<i>F. fusca</i>	<u>TATAAAAGAATATTTTTTTT--ATTTTATATAACTAATATTTATTTAACTGAAA</u>	90.9 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>TATAAAAGAATTTTTTTTTTTATTTTATATAACTAATATTTATTTAACTGAAA</u>	

trnN - trnS_I

Formicinae, Formicini	<i>F. fusca</i>	<u>TCCAGTTAATTATTTTTATAATCTTAATTTTAAATTATCCTGAAGTAATTA</u>	96.7 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>TCCAGTTAATTATTTTTATAATCTTAATCTTAAATTATCCTGAAGTAATTA</u>	

trnS_I - trnE

Formicinae, Formicini	<i>F. fusca</i>	<u>TTTTACTTCATAAATTATTTTATATTCATTATATTAAATTATATATTTATTTTTATTTATTTACTTTAATTTTTTTAAATAATTAATTTATATAG</u>	98.6 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>TTTTACTTCATAAATTATTTTATATTCATTATATTAAATTATATATTTATTTTTATTTATTTACTTTAATTTTTT-AAATAATTAATTTATATAG</u>	

trnF - ND5

Formicinae, Formicini	<i>F. fusca</i>	<u>CTATATAAATATAAATAAGTTAATTTAATTAAATCTTTCTTTATTTTCATTTATTATTTATAATTATAAATTTAATATCTATTAAA---AATTT-TTAAATAATT</u>	91.7 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>--ATATAAATATAAATAAGTTAATTTAATTAAATCTTTATTTATTTTAATTTATTATTTATAATTATAAATTTAATATCTATTAAATAAAATTTCTTAAATAATT</u>	

trnH - ND4

Formicinae, Formicini	<i>F. fusca</i>	<u>CTACTTAAACTCCTTAAAAATTAATATAAT</u>	100 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>CTACTTAAACTCCTTAAAAATTAATATAAT</u>	

ND4 - ND4L

Formicinae, Formicini	<i>F. fusca</i>	<u>TTATCATAATTAAATCTTAAAATTTATATAAAAAATTATTTAAAAATTTT</u>	100 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>TTATCATAATTAAATCTTAAAATTTATATAAAAAATTATTTAAAAATTTT</u>	

ND4L - trnT

Formicinae, Formicini	<i>F. fusca</i>	<u>TAATATATAT-----TAATAAATATTTCTTTATATTATACTATTATTAACCTTAATTATTAATATAATTAATATTACTTATATAG</u>	86.1 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>CAATATTAATATATACTAATAAATATTTCTTATATCATACTATTATTAACCTTAATTATTAAGATAATTAATATTACTTATATAG</u>	

trnT - trnP

Formicinae, Formicini	<i>F. fusca</i>	<u>ATTATAAGTAAAAA---TAATAAATTTTCAAAAAAAT</u>	83.3 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>ATTATAAGTAAAAAAAATAATAAATTTTCAAAAAAAT</u>	

<i>trnP</i> – <i>ND6</i>		<i>trnP</i>	<i>ND6</i>	
Formicinae, Formicini	<i>F. fusca</i>	<u>CTACTTTTTGAAATCAAATTAATTACTTTACTTTATGAATAAAG</u>		77.2 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>CTACTTTTTGAAATCAAATTAATTACT-----TATGAATAAAG</u>		
		*****	*****	
<i>ND6</i> - <i>CytB</i>		<i>ND6</i>	<i>CytB</i>	
Formicinae, Formicini	<i>F. fusca</i>	<u>AATTAATTAACTATCCCTAACCCCACTATATCCTCCTATACTAATAATTTATTATGAATAAAA</u>		86.0 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>AATTAATTAACTATCCCTAACCC--CTATGTC---CTATACTAATAATTTATTATGAATAAAA</u>		
		*****	*****	
<i>CytB</i> - <i>trnS₂</i>		<i>CytB</i>	<i>trnS₂</i>	
Formicinae, Formicini	<i>F. fusca</i>	<u>TTCTTCTTAAAATATATTTAAAAATTAAATTAAATTTTAAAATTTATTACTAAAATTATTTTTTTTATTTAGTTAATGAT</u>		100 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>TTCTTCTTAAAATATATTTAAAAATTAAATTAAATTTTAAAATTTATTACTAAAATTATTTTTTTTATTTAGTTAATGAT</u>		
		*****	*****	
<i>trnS₂</i> - <i>NDI</i>		<i>trnS₂</i>	<i>NDI</i>	
Formicinae, Formicini	<i>F. fusca</i>	<u>CTATTAACTTTTAAATTATAAATCAACTTAAATAATT</u>		93.3 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>CTATTAACTTTTAAATTACAAATCAACTTAAATAATT</u>		
		*****	*****	
<i>trnL₁</i> - <i>lrRNA</i>		<i>trnL₁</i>	<i>lrRNA</i>	
Formicinae, Formicini	<i>F. fusca</i>	<u>CCAAAGTAATTTTATTTTTACTTATTAATTAATATTTTTATTAATTTTTT-CTATTTTATACTTTATAATAATTATTTAATTTTTTTTAAAATAAATTATGTAACCTTAATAAATAAATAATTTAAATTTTTTAACTTTATTCCTTA</u>		96.0 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>CCAAAGTAATTTTATCTTTACTTATTAATTAATATTTTTATTAATTTTTTACTATTTTATACTTTATAATAATTATTTAATTTTTTT-AAAATAAATTATATAAATTAATAAATAAATAATTTAAATTTTTTAACTTTATTCCTTA</u>		
		*****	*****	
<i>lrRNA</i> - <i>trnV</i>		<i>lrRNA</i>	<i>trnV</i>	
Formicinae, Formicini	<i>F. fusca</i>	<u>TATAAAATTTTATTTAAAAGATATTTTTATTTTATTAATTAATAATTCAAAAAAAAA--TTAATAAATTAATTTATTAATTAAAAATTTTA</u>		97.0 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>TATAAAATTTTATTTAAAAGATATTTTTATTTTATTAATTAATAATTCAAAAAAAAAAATTAATAAATTAATTTATTAATTAAAAATTTTA</u>		
		*****	*****	
<i>trnV</i> - <i>srRNA</i>		<i>trnV</i>	<i>srRNA</i>	
Formicinae, Formicini	<i>F. fusca</i>	<u>CTAAAACTTTAATTTATTTATTATTATTATTAATTCTAAATA</u>		86.3 %
Formicinae, Formicini	<i>F. se/ysi</i>	<u>CTAAAACTTTAATTTATTTATTATTATT---AATTCTAAATA</u>		
		*****	*****	

<u>trnV</u> - <u>trnM</u>		<u>trnV</u>	<u>trnM</u>
Myrmicinae, Solenopsidini	<i>S. richteri</i>	<u>CTAAAATTTTTTCACATATACTAATAACTAATTCTTAATTATATATTTTAAATCTATAAAAATTTTTTAAAAATATAAGTAA</u>	
Myrmicinae, Solenopsidini	<i>S. geminata</i>	<u>CTAAAACTTTATCAATTATGTAAATAACTAACTCTTAATTATATATTTTAAATCCATAAAAATTTTT---AAATATAAGTAA</u>	81.9 %
Myrmicinae, Solenopsidini	<i>S. Invicta</i> (PMS)	<u>CTAAAATTTTTTCACATATACTAATAACTAATTCTTAATTATATATTTTAAATCTATAAAAATTTTTTAAAAATATAAGTAA</u>	
Myrmicinae, Solenopsidini	<i>S. Invicta</i> (VMS)	<u>CTAAAATTTTTTCACATATTCTAATAACTAACTCTTAATTATATATTTTAAATCCATAAAAATTTTT-AAAAATATAAGTAA</u>	
		***** **	*****
<u>trnM</u> - <u>trnQ</u>		<u>trnM</u>	<u>trnQ</u>
Myrmicinae, Solenopsidini	<i>S. richteri</i>	<u>ACTTTCATTATAACAAAATTATTTTATATATAA</u>	
Myrmicinae, Solenopsidini	<i>S. geminata</i>	<u>ACTTTCATTATTCAAAATTTATCTTATATTTAA</u>	69.2 %
Myrmicinae, Solenopsidini	<i>S. Invicta</i> (PMS)	<u>ACTTTCATTATAACAAAATTATTTTATATATAA</u>	
Myrmicinae, Solenopsidini	<i>S. Invicta</i> (VMS)	<u>ACTTTCATTATAACAAAATTTATTTTATATATAA</u>	
		*****	*** **
<u>trnQ</u> - <i>ND2</i>		<u>trnQ</u>	<i>ND2</i>
Myrmicinae, Solenopsidini	<i>S. richteri</i>	<u>TTAAAATATACTTAATATTATCATATATTTATATACTAAAAATTTTATTA--ATTAATATTAATTTTTTAAATTAATTAATT---TCCTTATATTATTATTATTTTTACCC</u>	
Myrmicinae, Solenopsidini	<i>S. geminata</i>	<u>TTAAAATATA--TAAAATTTT-----ATACTAAAAATTACATTATTATTAATATTAATTTTTTAAATTAATAATTATACTCTTATATTATTATTATTTTTACCT</u>	69.2 %
Myrmicinae, Solenopsidini	<i>S. Invicta</i> (PMS)	<u>TTAAAATATATTTTAATATTATTATATATTTATATACTAAAAATTTTATTA--ATTAATATTAATTTTTTAAATTAATTAATT---TCCTCATATTATTATTATTTTTACCC</u>	
Myrmicinae, Solenopsidini	<i>S. Invicta</i> (VMS)	<u>TTAAAATATATTTTAATATTATCATATATTCATATACTAAAAATTTTATTA--ATTAATATTAATTTTTTAAATTAATTAGTT---TCCTTATATTATTATTATTTTTACCC</u>	
		***** ** *	***** ** **
<u>trnP</u> - <i>ND6</i>		<u>trnP</u>	<i>ND6</i>
Myrmicinae, Solenopsidini	<i>S. richteri</i>	<u>CTATTTTTTGAAACCTAACTATACCAGCAT-AATCTAATT-ATCTATAATTTTATGATAAAAT</u>	
Myrmicinae, Solenopsidini	<i>S. geminata</i>	<u>CTATTTTTTGAAAATTAACTATCACACATTAATTCTATT-ATTTAT----TTATGATAAAAT</u>	46.5 %
Myrmicinae, Solenopsidini	<i>S. Invicta</i> (PMS)	<u>CTATTTTTTGAAAAATAACTATACCCACAT----CTCATT-ATTAATA-TTTTATGATAAAAT</u>	
Myrmicinae, Solenopsidini	<i>S. Invicta</i> (VMS)	<u>CTATTTTTTGAAAAC TAACCATACCCATAT-AATCTCATTATCTCTA-TTTTATGATAAAAT</u>	
		***** ** *	*****
<u>trnL_I</u> - <i>lrRNA</i>		<u>trnL_I</u>	<i>lrRNA</i>
Myrmicinae, Solenopsidini	<i>S. richteri</i>	<u>CAAAGTAATTTTTTAAA---TTATTTATAATTAAT</u>	
Myrmicinae, Solenopsidini	<i>S. geminata</i>	<u>CAAAGTAATTTTTTTAAATTTTATTATAACTAAA</u>	60.0 %
Myrmicinae, Solenopsidini	<i>S. Invicta</i> (PMS)	<u>CAAAGTAATTTCTAAAA---TTATTTATAATTAAT</u>	
Myrmicinae, Solenopsidini	<i>S. Invicta</i> (VMS)	<u>CAAAGTAATTTTTTAAA---TTATTTATAATTAAT</u>	
		***** *	*****